Essays on Improving Quality and Safety in Highly Regulated Industries

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERISTY OF MINNESOTA
BY

Sehwon Kang

IN PARTIAL FULFILMENT OF THE REQUIRMENTS
FOR THE DEGREE OF
DOCTORAL OF PHILOSOPHY

Advisor: Dr. Rachna Shah

August, 2019

Acknowledgement

I would like to express my deepest gratitude to my advisor, Rachna Shah, who has been crucial for completing my dissertation and guiding me in moving on to the next stage. I would also like to thank my committee members, Kevin Linderman, Susan Goldstein, and Bryan Dowd. I have benefited enormously from their guidance, support, and encouragement throughout the process.

I also thank my department and faculty members. Anant Mishra and KK Sinha gave me a great deal of counsel and advice. I also learned an incredible amount from seminars taught by Chris Nachtscheim, Karen Donohue, and Mili Mehrotra. I also thank Kartik Natarajan and Natalie Huang, who shared their insights and helpful comments. And, I thank my longest fellow Kedong and other colleagues including Alison, Amir, Amy, Anna, Dwai, Eric, Gautham, George, Jason, Jim, Pettis, Rick, Siwen, Suvrat, Ujjal, Vincent, and Yi.

This research would not have been possible without the advice from experts in the healthcare and pipeline industries. Jill Ryan, Ashley Krueger, Stephen Schmaltz, Tetyana Shippee, Pinar Mandic, Blaine Keener, Robert Clarillos, and many others helped me to shape my research in important ways.

In addition, I would like to thank my family for their support during the process. I thank my father who is an unwearied mathematician inspiring me to become a researcher. I am particularly indebted to my brother-in-law, who has been a committed copy editor. Any errors and omissions are my responsibility.

Dedication

To my parents and sister,

for their unwavering support and encouragement.

Abstract

Managing quality and safety is critical in highly regulated industries because failing to do so can lead to serious negative consequences. One way to improve quality and safety is enhancing organizational focus, emphasis on a specific set of actions. To study various contexts of focus, I select three settings in highly regulated industries: acute-care hospitals, nursing homes, and oil pipeline operators.

First, I study *internally driven focus* as disproportionate emphasis on a medical specialty in acute-care hospitals. I examine the effect of focus strategy and its combined effects with patient experience practices, on quality performance measured as readmission rates and patient satisfaction. Using secondary data from 3,027 hospitals, I find that focus has undesirable effects on both measures. However, patient experience mitigates the negative influence of focus on readmission rates. I also find that an imbalance between focus and patient experience results in poor performance. There is no single magic bullet to improve the two performance measures.

Second, I study *externally driven change in attentional focus* where recurring visits are unannounced while initial visits are announced in advance at nursing homes. Drawing on the attention-based view, I examine the effects of announced and unannounced inspections on the immediate and sustained quality performance. Using a dataset from accredited nursing homes, I show that unannounced inspection visits lead to a more sustained increase in quality performance than announced visits. Thus, announcing the inspection in advance results in short-term gains but long-term disadvantages.

Finally, I study *externally driven focus* on a safety management program in oil transportation. The program requires pipeline operators to prioritize their resources to reduce incidents in high consequence areas (HCAs). I examine the effects of pipeline system complexity and the learning experience with the program, on safety performance measured as future incident cost. Using a panel dataset of 642 pipeline operators, I find that complexity increases the cost but organizational learning reduces it. Interestingly, complexity heightens the negative relationship between the experience and future incident cost. The program is fruitful for incidents in high consequence areas (HCAs), but not in non-HCAs, which substantiates the intent of the program.

Table of Contents

List of Tables

List of Figures

Chapter 1

Dissertation Overview

Operations management scholars and practitioners frequently deal with quality and safety problems and consider them as top priorities (Ball et al., 2017; Das et al., 2008; Dean & Bowen, 1994; Flynn et al., 1994; Hendricks & Singhal, 2001; Sousa & Voss, 2002; Su & Linderman, 2016; Zhang & Xia, 2013). However, managing quality and safety is more critical in highly regulated industries because failing to do so can lead to serious negative repercussions, such as injuries, fatalities, property loss, and environmental damage (TRB & NASEM, 2017). For instance, a single incident (e.g., the Deepwater Horizon oil spill, the Kalamazoo River oil spill, Fukushima Daiichi nuclear disaster) costed more than several billion dollars. Highly regulated industries are the industries that have a greater number of restrictions in federal regulations. According to the McLaughlin-Sherouse list, heavily regulated industries such as the healthcare, oil and gas, and air and water transportation industries, contain more than 10,000 restrictions in federal regulations, which are more than ten times greater than the overall mean (Al-Ubaydli & McLaughlin, 2014).

One way to improve quality and safety in highly regulated industries is through enhancing organizational focus, emphasis on a specific set of actions or activities (McDermott & Stock, 2011). The central tenet underlying focus approach revolves more efficient and effective use of resources, which facilitates improved quality and safety performance. The benefits of focus have been demonstrated in diverse contexts in highly regulated industries such as airlines and healthcare industries (Clark & Huckman, 2012; Huckman & Zinner, 2008; Hyer et al., 2009; KC & Terwiesch, 2011; McDermott & Stock, 2011; Tsikriktsis, 2007). Although the literature on focus approach has grown significantly in several decades, previous studies take for granted that organizational focus is internally driven. Therefore, an important question remains open whether *externally driven focus* also improve quality and safety performance. To fill this research gap, my dissertation, *Essays on Improving Quality and Safety in Highly Regulated Industries*, explores focus that are internally and externally driven. To study two aspects of focus, I select three settings in highly regulated industries: acute-care hospitals, nursing homes, and oil and gas pipeline operators in the United States. Specifically, I explore 1) internally driven disproportionate emphasis (focus) on a medical specialty in acute-care hospitals, 2) externally driven attentional focus on inspection visits in nursing homes, and 3) externally driven focus on high consequence areas in a safety management program of pipeline operators.

In chapter two, *"The Dark Side of Focus: Is Patient Experience the Cure,"* I study *internally driven focus* as disproportionate emphasis on a medical specialty in acute-care hospitals, and its effect on quality performance. In addition to focus approach, which conserves resources via disproportionate emphasis, I consider another improvement approach, patient experience practice, which expends resources to provide positive patient experiences. Because both approaches put conflicting demands on hospital management, I address the following question in this chapter: how do hospitals' execution of focus strategy and patient experience practices affect their performance? These dual goals pose a considerable challenge for hospital administrators because pursuing focus rests on variation reduction, while improving patient experience increases variation in delivery processes.

To study these questions, I examine both direct and indirect effects of focus approach and patient experience practices on quality performance measured as readmission rates and patient satisfaction. Specifically, I model indirect effects in two ways: One, "moderation" to evaluate the combined effects of the two approaches, and two, "matching" to examine the impact of the imbalance between them. Using secondary data from 3,027 hospitals over six years and econometric analyses, I find that focus and patient experience have opposing direct effects on both the performance measures. Focus has undesirable effects on performance that it predicts an increase in readmission rates and a reduction in patient satisfaction, which are negative outcomes from management and patient perspectives. In contrast, patient experience predicts a reduction in readmission and an improvement in patient satisfaction, which are positive and desirable outcomes. However, I also find that interesting results from indirect effects. First, the joint effect of focus and patient experience reduces readmission rates. Second, an imbalance between focus and patient experience results in poor performance, particularly when hospitals emphasize focus over patient experience. As a set, this chapter highlights the challenges of hospital administrators that would pursue multiple objectives. Focus approach may not work effective in certain context and there is no single magic bullet to improve the two performance measures.

Supply chain partners, regulatory and accrediting agencies often rely on inspections to help manage quality and safety. Two dominant strategies related to inspections are announcing the inspection visit in advance or making an unannounced visit with little advance notification. In chapter three, *"Does Announcing the Visit Matter? An Empirical Examination in US Nursing Homes,"* I study *externally driven change in attentional focus* on inspection visits in nursing homes when recurring visits are unannounced while an initial visit is announced in advance. In this context, this study addresses the following question: Do announced and unannounced inspections lead to an immediate and/or a sustained increase in overall quality performance? The attention-based view of the firm provides theoretical grounding to investigate this question by understanding the differing roles of the two inspections strategies. The theory implies that what firms do depends on what they focus their attention on (Ocasio, 1997; Shepherd et al., 2017). Drawing on this theory, I argue that announced inspection results in transient attention to the inspection standards while unannounced inspection results in sustained attention.

To answer this question, I examine the contrasting effects of the two inspections on quality performance using a four-year panel dataset on over 300 accredited nursing homes and difference-in-difference relative time-models. I find that both types of inspection visits increase quality performance. However, unannounced inspection visits lead to a more sustained increase in quality performance when compared to the announced visits. Thus, announcing the inspection in advance lead to short-term gains but results in long-term disadvantages. To the best of my knowledge, this is the first empirical research to investigate the sustained effect of announced and unannounced inspections on operational performance. Thus, the results have important implications to supply chain partners and inspection agencies, who need to choose between announced and unannounced inspections. Specifically, unannounced inspections are more effective where high sustained quality performance is critically important, such as healthcare organizations.

In chapter four, *"Organizational Learning, Complexity, and Safety Management Performance: Evidence from the Oil and Gas Transportation,"* I study *externally driven focus* of a safety management program that pipeline operators are required to prioritize their resources to reduce spill events in high consequence areas (HCAs). Pipeline operators should focus on HCAs because most damages were in HCAs when pipeline incidents occurred (Kowalewski, 2013). In this study, I address the following questions: 1) What is the relationship between complexity and pipeline safety performance, 2) What is the relationship between organizational learning through an HCA-focused program and pipeline safety performance, and 3) When does organizational learning become more effective?

To study these questions, I examine whether complexity of the pipeline system and learning experience from the program have impacts on subsequent incident cost. I also explore the moderating role of complexity on the relationship between learning experience and subsequent incident cost. Using a fourteen-year panel dataset of 642 pipeline operators and employing multiple econometric models, I find that while pipeline complexity increases subsequent incident cost, the experience with the safety management program reduces it. More interestingly, complexity heightens the negative relationship between the experience with the program and future incident cost. The results also show that learning experience reduces future incident costs in HCAs but not in non-HCAs, substantiating the intent of the program. As a set, I highlight the effectiveness of using an externally mandated, but internally developed safety management programs in oil and gas transportation. This study contributes to organizational learning literature, especially in highly regulated industries with high-hazard. The findings are also of practical importance to both pipeline operators and federal regulators.

Taken together, the three essays of my dissertation provide a deeper understanding of improving quality and safety in highly regulated industries through the lens of focus. I demonstrate that *internally driven focus* may not improve quality performance when pursuing multiple objectives and that *externally driven focus* improves quality and safety performance in highly regulated industries.

Table 1.1 Dissertation Overview

Essay	Research Questions (Focus)	Industry	Key (Independent; Dependent) Variables
1	Do focus and patient experience have combined effects on quality performance? (Focus as disproportionate emphasis)†	Hospitals	Focus and Patient Experience; Quality
2	Does announcing the inspection visit influence quality performance? (Attentional focus on inspections)‡	Nursing Homes	Inspection Announcement; Quality
3	Do organizations learn from safety management program? (Focus on high consequence areas)‡	Oil-Gas Pipelines	Safety Management Program; Safety

†: *internally driven focus*; ‡: *externally driven focus*

Chapter 2

The Dark Side of Focus: Is Patient Experience the Cure?

2.1 Introduction

Hospitals are under tremendous pressure to concurrently improve measures of clinical performance, such as readmission and mortality rates, as well as measures of patient satisfaction. Both types of outcomes have acquired added significance for hospital administrators in recent years because of new legislative and reimbursement policies. For instance, the Centers for Medicare & Medicaid Services (CMS) monitors hospitals and provides incentives for improving both types of performance through various initiatives, including the Hospital Value-Based Purchasing (VBP) Program and the Hospital Readmission Reduction Program (HRRP). However, improving both measures of performance simultaneously can put hospital administrators in a bind because the two measures require different approaches.

To improve clinical performance, such as readmission and mortality rates, hospitals frequently emphasize conformance initiatives and adopt standardized clinical protocols and administrative practices to reduce rework and waste, among other approaches (Cook et al., 2014; Senot et al., 2016b). One of these “variance-reduction” approaches is adopting a “focus” strategy, which emphasizes growing specific healthcare specialties such as cardiology and hip replacement. The focus strategy has been shown to improve clinical and cost performance (Lee et al., 2015; McDermott & Stock, 2011). Such management decisions of where and how much to focus are mostly discretionary and illustrate how the organization rationalizes and allocates its resources.

Regulatory agencies also influence hospitals to implement practices aimed towards improving overall patient satisfaction through financial incentives, penalties and regulatory policies. Such tactical initiatives require and consume hospital resources, and may increase variance in delivery processes. For instance, creating a positive patient experience may require “amenities of care” (Donabedian, 1988) such as quietness and cleanliness, as well as time-consuming interaction between hospital caregivers and patients (Chandrasekaran et al., 2012; Donabedian, 1988; Senot et al., 2016b; Westbrook et al., 2014). Hospitals may incur significant staffing and training costs to allow caregivers to spend extra time with patients to determine and fulfill their specific needs (Senot et al., 2016b). Interestingly, the core of the VBP Program (CMS, 2014) is the Patient Experience of Care initiative, a reimbursement policy requiring hospitals to show acceptable performance on patient experience measures. Failure to show acceptable performance, by demonstrating a hospital's improvement over time and in comparison to other hospitals, can result in penalties.

Improving these two distinct outcomes (i.e. clinical performance and overall patient satisfaction) by pursuing two improvement approaches - via focus and patient experience - puts conflicting demands on hospital management: *conserving* resources via disproportionate emphasis on a medical specialty versus *expending* resources to provide positive patient experiences. These dual goals also pose a considerable challenge for hospital administrators because pursuing focus rests on variation reduction, while improving patient experience increases variation in delivery processes. Our extensive review of the related literature shows that both focus, a "variation-reduction strategy," and patient experience, based on "variation-increasing tactics," have been studied extensively. However, none of the existing studies have examined them together, nor do they assess their individual or combined impact on these two distinct performance outcomes – clinical performance and overall patient satisfaction. Our study addresses this gap and inquires: how does a hospital's execution of a focus strategy and patient experience practices, two potentially conflicting approaches, impact its performance?

Following previous research, we measure focus as a disproportionate emphasis on a medical specialty (KC & Terwiesch, 2011; McDermott & Stock, 2011). We select the cardiology specialty as our setting for focus because it has been studied extensively, accounts for a significant proportion of hospital revenues and patient volume, and its effects on performance are fairly well understood (Ding, 2015; McDermott & Stock, 2011). Similar to previous studies, we conceptualize patient experience as patient perception of care, and use HCAHPS data to measure it (Donabedian, 1988; Nair et al., 2013; Senot et al., 2016b). We measure hospital performance as readmission rates and overall patient satisfaction, both of which have been broadly used in previous literature (Boulding et al., 2011; Cook et al., 2014; Ding, 2015; Marley et al., 2004). In this study, we examine both the direct and indirect effects of focus and patient experience on readmission and overall patient satisfaction. We model indirect effects in two ways. First, we use "moderation" to evaluate the combined effects of the two approaches. Second, we use "matching" to examine the impact of the imbalance between them.

We address our research question using data compiled from multiple secondary sources for 3,027 U.S. hospitals over a 6-year period. We test our hypotheses using seemingly unrelated regression (SUR) analyses. Our results show that focus and patient experience have opposing direct effects on the two measures of performance. Focus predicts an increase in the readmission rates and a reduction in patient satisfaction, i.e. negative outcomes from management and patient perspectives. In contrast, patient experience predicts a reduction in readmission and an improvement in patient satisfaction, i.e. positive and desirable outcomes. The indirect effects are more nuanced, but equally interesting. First, we find that the joint effect of focus and patient experience reduces the readmission rates but has no significant impact on patient satisfaction. Second, an imbalance between focus and patient experience increases readmission and decreases patient satisfaction, both undesirable outcomes from management and patient perspectives. We conduct robustness checks to substantiate these results and post hoc analyses to explore these relationships. In doing so, we find that the undesirable influence of focus on readmission and patient satisfaction is greater for hospitals with higher bed occupancy rates and teaching responsibilities; we find no such dependence for patient experience. Our investigation into imbalance effects shows that its negative effects on performance are exacerbated when a hospital emphasizes the focus strategy more than patient experience practices. As a set, our results highlight the challenges hospital administrators face in pursuing multiple objectives. We show that while there is no single magic bullet to improve both clinical performance and patient satisfaction, a balanced approach can be effective for achieving multiple objectives.

2.2 Literature Review, Theoretical Grounding and Hypotheses

There are numerous studies examining hospital performance (Andritsos & Tang, 2014; Bechel et al., 2000; Hyer et al., 2009; KC & Terwiesch, 2011; Marley et al., 2004; McDermott & Stock, 2011; Nair et al., 2013; Senot et al., 2016b). However, it is difficult to compare their results because they vary greatly in their independent and dependent variables. Therefore, to understand the current state of knowledge and draw useful insights, we concentrate our literature review on the healthcare delivery context related to: 1) focus; 2) patient experience; and 3) performance measures. We summarize the most relevant studies in Table 2.1. Although these studies differ in terminology, measures of performance, and the unit of analysis, they allow us to make a few overarching observations about this literature. First, a majority of studies are conducted in the cardiology setting. Second, while focus is measured in a variety of ways, the most common measure is “case focus” (defined as “cardiology cases as a proportion of total cases”). Third, performance measures are most frequently either cost-related (e.g., length of stay, cost per day, cost per admission) or mortality, while readmission rates and patient satisfaction are included less frequently.

Table 2.1 Research on Focus in the Healthcare Sector

Article	Research Question	Measure of focus (Unit of analysis)[a]	H[b]	P[b]	Dependent Variables	Control Variables	Results
Lee et al. (2015)	Effect of focus and vol on outcome	(135 Hospital): Heart disease Patients / Total Patients	X		Risk adjusted mortality	Pat-Hosp Characteristics; Pat. Vol	- Focus reduces mortality. - Volume is not.
Ding (2015)	Effects of focus and quality improvement initiatives on clinical quality	(210 Hospital): Cardiology Procedures / Total Procedures; (Department): Heart attack Procedures / Cardiology Procedures	X		Risk adjusted mortality; Risk adjusted readmission; Process care	Hosp. Characteristics; Procedure Volume	- Department focus improves mortality but not readmission; - Hosp focus does not impact performance.
Andritsos & Tang (2014)	Moderating effect of focus on process quality and resource use	(1298 Patients, Operating Unit): Concentration of Cardiac Patients across all different conditions		X	Resource usage; Length of stay	Hosp. Characteristics	Resource usage reduction on process quality is higher for low focus than greater focus hospitals.
Cook et al. (2014)	Adopting a focus factory model within a solution shop	(Treatment): Focused-factory adoption			Length of Stay; Mortality; Cost; Readmission	Pat. Characteristics	- Focus reduces resource use & cost; no improvement in clinical outcome.
Ding (2014)	Effects of focus, experience, & ownership on productive efficiency	(3700 Hospital): Concentration of clinical area across all areas	X		Operating cost	Hosp. Characteristics; Labor cost; Legislation	- Focus & experience improve productive efficiency.
Clark & Huckman (2012)	Impact of focus on mortality	(382 Hospital): Cardiovascular Patients / Total Patients	X		Risk adjusted mortality	Hosp. Characteristics; Pat. Volume	Focus reduces mortality.
KC & Terwiesch (2011)	Effect of focus (at three levels) on operational performance	(Hosp.): Cardiac Pat. / Total Pat.; (Operating Unit): Revascular Pat. / Cardiac Pat.; (Process): CABG Pat. / Revascular Pat.	X	X	Risk adjusted mortality; Risk adjusted length of stay	Pat-Hosp Characteristics; Total Pat., Cardiac Pat. Volume	- At hosp. level, focus lowers mortality & LOS, but effect is not significant with IV estimation; - At operating unit level, focus lowers LOS even with IV estimation.
McDermott & Stock (2011)	Effect of focus on cost performance	(Cardiology; NY Hospitals): Cardiac Cases / Total cases; Cardiac Pat-days / Total Pat-days; Coronary Care Beds / Total Beds	X		Total cost; Cost per day	Pat-Hosp Characteristics	All three measures of focus reduce both measures of cost

a: Unit of analysis in parenthesis; b: H – Hospital level, P – Patient level

Next, we review the individual and joint effects of focus and patient experience on performance. In studies related to focus, we find that, in general, greater focus reduces most measures of cost (Cook et al., 2014; McDermott & Stock, 2011). However, its impact on readmission rates is not significant (Ding, 2015), and the evidence of its impact on mortality is mixed. Some studies identify no significant impact of focus on mortality (KC & Terwiesch, 2011), while others show that it reduces mortality (Clark & Huckman, 2012; Lee et al., 2015). We find no studies linking focus with patient satisfaction. In terms of patient experience, there are only a handful of studies examining its performance impact. Senot et al. (2016b) show that patient experience, measured as experiential quality, reduces readmission rates. While the relationship between patient experience and patient satisfaction is logical and should be widespread, Chandrasekaran et al. (2012) suggest that the effect is neither universal nor strong. It is noteworthy to point that we do not find any studies, which examine the impact of both focus and patient experience on readmission rates and patient satisfaction, as we do in the current study.

Focus

There is unambiguous evidence that focus is associated with improved performance in both manufacturing and service settings. For instance, in manufacturing, focus is associated with higher productivity (Brush & Karnani, 1996) and higher firm market value (Wernerfelt & Montgomery, 1988). In services, the performance benefits of focus have been demonstrated in diverse contexts such as airlines (Tsikriktsis, 2007), banking (Staats & Gino, 2012), and healthcare (Clark & Huckman, 2012; Huckman & Zinner, 2008; Hyer et al., 2009; KC & Terwiesch, 2011; McDermott & Stock, 2011). Researchers argue that performance is improved because focusing on a narrow range of consistent and related tasks allows firms to use limited organizational resources more effectively. Additionally, by emphasizing work (customers) of a particular type, firms develop greater expertise in the associated tasks, and thus, appropriate advantages due to learning effects.

The central tenet underlying each of the above mechanisms revolves around variance-reduction and more efficient use of resources, which facilitate improved performance. These same mechanisms apply in the healthcare context. Focusing on a particular set of clinical conditions reduces variation among patients, and permits nurses and doctors to more readily apply their expertise and experience, making it easier to coordinate activities and communicate relevant information (Shortell et al., 1994). Increased coordination and communication, along with greater learning gained from focusing on high-volume conditions, should result in immediate improvement in clinical performance and lower future failure rates, such as readmission. Since high readmission rates signal poor clinical quality, lower readmission rates is considered good performance. Thus, we posit that:

Hypothesis 1a. Focus is negatively associated with (i.e. reduces) readmission rates.

Emphasizing a particular line of service or a narrow set of conditions implies that a hospital is likely to dedicate resources (such as equipment and space) into focal areas by redirecting them from other areas, perhaps to the detriment of the non-focal areas. Thus, while the increased learning and expertise gained from focus may medically benefit patients in the focal area, the emphasis on variance reduction is at odds with patient experience. This could result in lower overall satisfaction for the patients in the entire hospital. This dynamic is likely to be more accentuated in an acute care hospital setting, with a more diverse patient population compared to a specialty hospital. Thus, we propose the following hypothesis:

Hypothesis 1b. Focus is negatively associated with (i.e. reduces) overall patient satisfaction.

Patient Experience

In recent years, patient experience has attracted considerable attention from practitioners. For instance, a report by the Institute of Medicine (IOM 2001, 2009), highlights the importance of patient experience as one of the key elements of high-quality care. In the same vein, the Agency for Healthcare Research and Quality (AHRQ) and CMS regularly survey patients on their hospital experience (IOM 2009)

Despite its salience, patient experience has not been well investigated in the management literature. A few noteworthy exceptions include Chandrasekaran et al. (2012), Nair et al. (2013), and Senot et al. (2016a, 2016b), each invoking patient experience in their conceptualization of experiential quality. The two concepts are closely related; while similarities between them abound, the differences are fine-grained and nuanced. For instance, patient experience entails both humanistic factors such as how services are delivered to the patient, and mechanistic factors such as the physical settings where services are provided (Donabedian, 1988). In contrast, experiential quality focuses on the humanistic aspects of caregiver-patient interactions, such as interpersonal communication and the speed of addressing patient needs (Chandrasekaran et al., 2012; Senot et al., 2016a, 2016b). Nevertheless, both patient experience and experiential quality center on incorporating and responding to unique patient needs. This enhanced attention to patients should result in identifying and resolving clinical issues as they occur during the delivery of care, reducing future readmission as well as enhancing overall patient satisfaction. Existing empirical evidence, albeit limited, supports these relationships. For instance, a positive patient experience is shown to reduce average length of stay and readmission rates (Boulding et al., 2011; Nair et al., 2013; Senot et al., 2016b). Given the evidence, we expect:

> *Hypothesis 2a. Patient experience is negatively associated with (i.e. reduces) readmission rates.*

> *Hypothesis 2b. Patient experience is positively associated with (i.e. increases) overall patient satisfaction.*

Indirect Effects

In hypothesizing the direct relationships between focus, patient experience, and the two measures of performance, we posit that focus helps reduce variance and employ resources more efficiently, whereas improving patient experience induces variances which may require firms to expend resources. Undoubtedly, management is faced with tough choices in their effort to balance these two approaches. In such cases, management can either pursue both approaches in the hopes of amplifying their individual effects or mitigating the negative effects of one with the other. In contrast, management may favor one approach over the other. Organizational theorists have popularized these two perspectives under the general umbrella of "fit," and specifically refer to them as "moderation" and "matching" (Venkatraman, 1989). Following this conceptualization, we analyze the indirect effects of focus and patient experience as moderation and matching in this study. The moderation perspective implies that the predictor (independent variable) and the moderator (second independent variable) jointly determine the impact on the criterion (dependent) variable (Venkatraman, 1989). In contrast, the matching perspective is invoked when researchers suggest that a balance (or imbalance) between two variables impacts performance. In the section below, we describe how these apply in our particular context.

Indirect Effects as Moderation

We previously argue that both focus and patient experience individually result in lower readmission rates. If hospital management pursues both approaches simultaneously, their combined effects may have a different or greater effect than either of them individually. When hospitals emphasize one specialty relative to others, they are able to underscore a narrower, cohesive set of tasks. This allows them to appropriate greater learning about the medical condition as well as to identify more and better improvement opportunities in the care delivery process. In a focused setting, experienced professionals can tailor their actions to meet the needs of the patients within that focused set. Thus, communication between providers and patients can be more targeted and instructions more specific, resulting in better understanding by patients at discharge, greater adherence to instructions, and thus, lower future readmission rates. The physical environment of a focal area can also reduce readmission rates: Cleaner rooms improve hygiene and prevent hospital-acquired infections, and quieter rooms promote healing (Banerjee, 2017). Together, focus and patient experience are likely to have a greater effect in reducing readmission rates than their individual effects, resulting in the following hypothesis:

Hypothesis 3. Focus and patient experience have a negative (improvement) joint effect on readmission rates.

In predicting overall patient satisfaction above, we hypothesized that it is negatively associated with focus in H1b, and positively associated with patient experience in H2b. Thus, on the one hand, enhancing patient experience may impact patient satisfaction in the entire hospital, and its positive effect may help overcome the negative effect of focus. On the other hand, patient experience may not be sufficiently strong to compensate for the negative effect of focus on overall patient satisfaction. In light of the ambiguous theoretical arguments, we believe that, together, patient experience and focus are likely to have a net positive effect on overall patient satisfaction.

Hypothesis 4. Focus and patient experience have a positive joint effect on overall patient satisfaction.

Indirect Effects as Matching

Favoring either focus or patient experience over the other may result in lower performance. Instead, researchers have shown that, in comparison, a balanced approach typically results in better performance (He & Wong, 2004). Organization theorists assert that firms must carefully manage the tradeoffs when facing two conflicting approaches which create tension (Birkinshaw & Gupta, 2013). They conclude that firms capable of balancing the two approaches simultaneously are likely to achieve superior performance relative to firms emphasizing one at the expense of the other. Applying this thinking to our context, we argue that when a hospital pursues focus without concurrent attention to patient experience, performance may suffer. Focus may dominate daily practices, but providers may not be sufficiently trained, staffed, or motivated to consider patients' specific needs, thus influencing readmission rates. The reverse may also be true. To enhance patient experience, if hospital administrators relentlessly cater to patient needs while ignoring or sacrificing the process standards or clinical protocols more closely aligned with a focused approach, clinical outcomes may suffer. Such an imbalance may result in higher readmission rates, suggesting the following hypothesis.

> *Hypothesis 5. Relative imbalance between focus and patient experience is positively associated with (i.e. increases) readmission rates.*

We expect that a similar dynamic may impact patient satisfaction as well. An overemphasis on either focus or patient experience, without concurrent attention to the other, will have a negative impact on overall patient satisfaction, suggesting the following hypothesis.

> *Hypothesis 6. Relative imbalance between focus and patient experience is negatively associated with (i.e. reduces) overall patient satisfaction.*

2.3 Research Design and Data

We investigate the direct and indirect effects of focus and patient experience on performance in U.S. acute care hospitals. We exclude critical access hospitals (a designation for some rural hospitals in underserved areas) and Veterans Administration hospitals because they do not report certain necessary data to the Centers for Medicare & Medicaid Services (CMS). The final sample is an unbalanced panel of 3,027 acute care hospitals that provide heart failure care.[1] We select heart failure because services for this diagnosis are broadly provided, and such selection helps reduce contextual and spurious effects that might result from including multiple disease groups.[2] Heart failure care represents a high-volume and high-revenue service set. About 5.7 million adults have heart failure, causing about one in nine deaths in the U.S. in 2013 (Mozaffarian et al., 2016). The direct medical costs of treating heart failure were estimated to be $32.4 billion in 2015 (Heidenreich et al., 2011).

To test our hypotheses, we use data from multiple CMS reports. All variables and corresponding data sources are listed in Table 2.2. The CMS cost report provides information on total beds, bed days and hospital location. We obtain teaching status and hospital-level numbers of licensed practical nurses and registered nurses from the CMS Provider of Services file. Patient

[1] Heart failure is “a condition in which the heart can't pump enough blood to meet the body's needs” (National Heart, Lung, and Blood Institute, 2016)
[2] http://www.aha.org/research/rc/stat-studies/fast-facts2015.shtml (2013 AHA Annual Survey)

experience and patient satisfaction are from the CMS HCAHPS survey, and 30-day readmission rates and clinical process quality are from CMS Hospital Compare. We obtain case mix index from CMS Impact files. We merge the data using CMS's unique hospital identifiers. Because all data sources contain information at the hospital level, our unit of analysis is the acute care hospital.

Our study period spans July 2007 to June 2013 for two reasons. First, the CMS Hospital Compare provides three year rolling averages with the period beginning in July of each year. Second, the HCAHPS reports began providing data on a one-year rolling average basis in October 2006, even though the records are updated quarterly.[3] Because the readmission rates of CMS are available only for three-year rolling periods, our dataset has two time periods: July 2007 – June 2010 and July 2010 – June 2013. We do not include data after June 2013 because the VBP and Hospital Readmission Reduction Programs initiated hospital performance incentives in late 2013, which is likely to influence our substantive variables. Therefore, we only analyze data prior to the implementation of these incentives. We aggregate all variables to three year rolling averages for our analyses. For robustness checks, we create a new dataset by allowing partial overlapping of time-periods (July 2007 – June 2010, July 2008 – June 2011, July 2009 – June 2012, and July 2010 – June 2013). This extended dataset has four time-periods.

[3] http://www.hcahpsonline.org/files/HCAHPS Fact Sheet, revised1, 3-31-09.pdf

Table 2.2 Summary Statistics and Sources

Variable	Description (n=3,027, Observations=5,862)	Mean	SD	Min	Max	Source (CMS)
Performance Measures						
Readmission rate (%)	Heart failure (HF)	23.856	2.206	16.6	33.8	Hospital Compare
Patient satisfaction	overall (logit transformed)	0.670	0.370	-0.911	2.876	HCAHPS survey
Independent Variables						
Focus	HF cases over total cases	0.242	0.100	0.005	1.000*	Medicare volume
Patient Experience	hospital average score (logit-ed)	0.808	0.235	-0.486	3.316	HCAHPS survey
Controls						
Time	time dummies					
*State***	state dummies (*k*=51)					Provider of Services
Ownership	non-profit (64.0%), profit (18.8%), Gov't (17.2%)					Provider of Services
Teaching	whether hospital has a medical school (Teaching: 33.1%)					Provider of Services
Location	whether hospital is located in urban area (Urban: 72.1%)					Cost report
Hospital size	number of beds	204.38	177.36	12	2127	Cost report
Bed occupancy rate	total bed days over total bed days available	0.570	0.176	0.090	1.085	Cost report
Case Mix Index		1.421	0.271	0.784	3.234	Impact files
Nursing intensity (*LPNs*)	number of LPN/LVNs over number of LPN/LVNs and RNs	0.143	0.140	0	1.000	Provider of Services
Clinical Process Quality$_{HF}$	Logit-ed score from heart failure	4.015	2.037	-5	9.210	Hospital Compare
Clinical Process Quality$_{all}$	Logit-ed score from heart attack, heart failure, and pneumonia	3.193	1.076	-1.754	9.210	Hospital Compare

* Three hospitals have max value of 1.000. The second highest value is 0.789. ** We control for state effect in specified models. k includes Washington D.C.

Dependent Variable: Performance

Readmission Rates in Heart Failure. Hospital readmission rates are an important measure of clinical performance (Axon & Williams, 2011). We use readmission rates for heart failure diagnosis from the CMS Hospital Compare file. The CMS data include Medicare fee-for-service (FFS) patients aged 65 or older when such patients are readmitted for the same diagnosis within 30 days after discharge. CMS reports a risk-adjusted measure because it is important in measuring and interpreting clinical outcomes. CMS creates this measure for each hospital using hierarchical logistic regression adjusting for patient age, gender, and comorbidities. The value of risk-adjusted readmission rates is the number of "predicted" outcomes over the number of "expected" outcomes, multiplied by the national readmission rates (CMS, 2014). We exclude hospitals whose readmission rates are constructed from fewer than 25 patients, following the CMS guideline.

Patient Satisfaction. Patient satisfaction is another important dimension in assessing hospital performance (Kane et al., 1997; Marley et al., 2004). This is an outcome measure because it is assessed after a clinical treatment has ended (Kane et al., 1997).[4] Patient satisfaction is defined as "how the patients judge the overall hospital experience" (Marley et al., 2004). We use data from an HCAHPS questionnaire which asks how inpatients rate their hospital overall, from 0 ("worst hospital possible") to 10 ("best hospital possible"). The HCAHPS data provide the aggregate percent of inpatients in three clustered answer categories for each hospital, scores of "6 or lower," "7 or 8," and "9 or 10." We create a patient satisfaction measure following (Chandrasekaran et al., 2012): multiplying the percentage of responses in each category by -1, 0, and 1, respectively, and then summing them. The reported data on the patient satisfaction measure include all hospital patients (CMS, 2013).

Independent Variables: Direct Effect

Focus. Previous healthcare research on focus has been conducted primarily in the cardiac care context, where most researchers have employed a volume-based measure of focus using patient-level, case-level, or procedure-level data (Clark & Huckman, 2012; KC & Terwiesch, 2011; Lee et al., 2015). For example, McDermott and Stock (2011) measure focus as a ratio of the number of cardiac cases relative to total cases. Ding (2015) uses both hospital-level focus measured as cardiology procedures relative to total procedures, and department-level focus measured as heart attack procedures relative to all cardiology procedures. Similarly, KC and Terwiesch (2011) measure three levels of focus: firm, operating unit, and process. Other notable researchers have

[4] CMS collects HCAHPS survey from inpatients between 48 hours and six weeks after their discharge.

used Herfindahl-Hirschman index as an alternative volume-based measure (Andritsos & Tang, 2014; Ding, 2014). We follow the more frequently used volume-based measures, and compute focus ($Focus_{it}$) as the ratio of number of heart failure cases relative to the total number of cases at hospital i and during time t.

Patient Experience. Patient experience incorporates patient perception of healthcare service delivery, including care and non-care aspects (Chandrasekaran et al., 2012; Donabedian, 1988; Manary et al., 2013). Care includes interactions between the patient and medical providers, such as general communication characteristics, pain management, and responsiveness to patient needs (Chandrasekaran et al., 2012; Nair et al., 2013; Senot et al., 2016b). Non-care captures perceptions of care amenities, including the hospital environment (e.g., quietness, cleanliness) (Donabedian, 1988; Epstein et al., 2010; Westbrook et al., 2014). We use HCAHPS survey items (see Appendix A, Table 1), a national initiative to consistently measure patient experience in the domains of communication, responsiveness, and hospital (physical) environment (Giordano et al. 2010). While previous studies do not consider the non-care aspects of patient experience (ENV 1, ENV 2), our study uses a more comprehensive measure of patient experience by including all domains in the HCAHPS survey (Chandrasekaran et al., 2012; Nair et al., 2013; Senot et al., 2016a, 2016b) (Chandrasekaran et al. 2012, Nair et al. 2013, Senot et al. 2016a, 2016b). We assess our patient experience measure for internal consistency, and find that Cronbach's alpha is 0.935.

CMS provides data for all hospitals which report HCAHPS responses from 100 or more patients. CMS addresses potential bias in the HCAHPS survey data by using their Patient-Mix Adjustment (PMA) model, which addresses the effect of patient characteristics on HCAHPS responses. The model includes education, self-rated health status, language at home, age, type of service (maternity, surgical, or medical), survey mode, and interaction effects of age and service line type.[5]

To create a comprehensive measure of patient experience ($PatExp_{it}$), we first extract the percentage of survey respondents who answered "Always" or "Yes" for each measurement item, and average these values across the eight items. Following previous studies, we use a logit transformation to satisfy the normality assumption (Chandrasekaran et al., 2012; Nair et al., 2013; Senot et al., 2016b).

[5] http://www.hcahpsonline.org/modeadjustment.aspx

Independent Variables: Indirect Effect

We test two types of indirect effects for focus and patient experience, "fit as moderation" and "fit as matching" (Venkatraman, 1989). Fit as moderation corresponds to the joint effects of the variables and is measured as the product of focus and patient experience (Cao et al., 2009; He & Wong, 2004). It implies that focus and patient experience are complements and add value in addition to their individual effects toward improving hospital performance. In contrast, fit as matching corresponds to the (im)balance between the two variables and is measured as the absolute value of the difference between focus and patient experience (Birkinshaw & Gupta, 2013; He & Wong, 2004). Fit as matching examines how the degree of alignment (a small or large absolute difference) between focus and patient experience affects hospital performance. We standardize all the continuous variables, and focus and patient experience before computing the interaction and absolute difference terms to avoid multicollinearity with the main effects.

Control Variables

We incorporate several key control variables from previous studies to control for hospital characteristics that may influence hospital performance.

Time. We include time dummy variables to account for the data collection periods. We consider July 2007 – June 2010 as the first time period and July 2010 – June 2013 as the second time period. This allows the models to capture changes over time in industry readmission rates and patient satisfaction.

Ownership. Ownership is important in predicting hospital reactions corresponding to differences in operating policy. Using two ownership dummy variables, we classify hospitals by three ownership groups: not-for-profit (base level), for-profit, and governmental hospitals. This classification is commonly used in previous studies (Andritsos & Tang, 2014; Ding, 2014).

Teaching. We control for teaching status because this may be related to hospital performance. Teaching status indicates the different mission and resource allocation of hospitals (Goldstein & Iossifova, 2012). We measure teaching status using a dummy variable, which indicates whether a hospital is affiliated with a medical school.

Location. Hospitals located in urban areas are coded as one and otherwise zero. This indicator is based on Core Based Statistical Areas.

***Beds* (*Hospital Size*)**. We control for hospital size measured as the total number of beds. We use a log transformation because of skewness.

Bed Occupancy Rates. Bed occupancy represents hospital utilization measured by the total patient days over the total hospital bed days available (McDermott & Stock, 2011).

***Case Mix Index* (*CMI*)**. According to CMS, case mix index represents a hospital's average diagnosis-related group (DRG) weight. It is calculated by total DRG weights for all Medicare discharges over total discharges. This reflects the extent of patient medical complexity.

Nursing Intensity. Previous studies show that the level of nurse staffing, especially registered nurses, is related to clinical performance (Mark et al., 2004). Registered nurses are associated with greater expertise than licensed (practical or vocational) nurses, and are likely to have a systematic influence on hospital performance. Nursing intensity is measured as the number of registered nurses relative to the total number of licensed and registered nurses, in a hospital.

***Clinical Process Quality* (*CPQ*).** Following Senot et al. (2016b), we use CPQ to capture important dimensions of conformance (e.g. whether appropriate medications or treatments are delivered to patients in a timely manner) and include it to control for endogeneity in predicting hospital readmission. Similar to previous researchers, we measure CPQ across the categories of heart failure (HF), acute myocardial infarction (AMI), and pneumonia (PN) from the process of care measures of CMS Hospital Compare Data (see Appendix A, Table 2). Following Chandrasekaran et al. (2012), Nair et al. (2013), and Senot et al. (2016b), we compute separate composite scores for HF measures ($cpq_{hf,it}$), and then combined for HF, AMI, and PN ($cpq_{all,it}$), by averaging across items and taking a logit transformation. CPQ for HF is used in the readmission rate models (Table 2.4), while the combined measure is used in the overall patient satisfaction models (Table 2.5).

Empirical Strategy

We estimate separate models to test the direct and indirect effects of focus and patient experience on readmission rates and overall patient satisfaction. The models include focus ($Focus_{it}$), patient experience ($PatExp_{it}$), and all controls ($\boldsymbol{X}_{it}$) of given hospital and time. We examine joint ($Focus_{it} \times PatExp_{it}$) and relative imbalance effects ($|Focus_{it} - PatExp_{it}|$) in separate models. To partial out non-linear effects from the joint effect models, we include quadratic terms for focus and patient experience. It is considered a conservative approach for testing joint effects (Dawson 2014, Ganzach 1997). We run two equations for each dependent variable, first with joint effects and then with imbalance effects. The following equations describe the four models.

$$Readmission\ Rate_{it} = \beta_0 + \beta_1 Focus_{it} + \beta_2 Focus_{it}^2 + \beta_3 PatExp_{it} + \beta_4 PatExp_{it}^2 + \beta_5 Focus_{it} * PatExp_{it} + \beta_6 C_{hf,it} + \boldsymbol{X}_{it}\gamma + \varepsilon_{it} \quad (1)$$

$$Readmission\ Rate_{it} = \beta_0 + \beta_1 Focus_{it} + \beta_2 PatExp_{it} + \beta_3 |Focus_{it} - PatExp_{it}| + \beta_4 C_{hf,it} + \boldsymbol{X}_{it}\gamma + \varepsilon_{it} \quad (2)$$

$$Satisfaction_{it} = \beta_0 + \beta_1 Focus_{it} + \beta_2 Focus_{it}^2 + \beta_3 PatExp_{it} + \beta_4 PatExp_{it}^2 + \beta_5 Focus_{it} * PatExp_{it} + \beta_6 C_{all,it} + \boldsymbol{X}_{it}\gamma + \varepsilon_{it} \quad (3)$$

$$Satisfaction_{it} = \beta_0 + \beta_1 Focus_{it} + \beta_2 PatExp_{it} + \beta_3 |Focus_{it} - PatExp_{it}| + \beta_4 C_{all,it} + \boldsymbol{X}_{it}\gamma + \varepsilon_{it} \quad (4)$$

Estimation of combined effects on readmission rates (1) and patient satisfaction (3) are interdependent because equations sharing observable characteristics might also share unobservable characteristics. This may lead to correlation between their residual terms (Moulton, 1990). Seemingly Unrelated Regression (SUR) is used when a system of equations are related due to the correlation in residual terms, although the equations seem unrelated (Devaraj et al. 2004, Zellner 1962). In such a situation, SUR is appropriate because it produces smaller standard errors even when the residual terms of the equations are highly correlated. SUR results would be identical to the results from OLS regression without any added efficiency (i.e. smaller standard errors) if the right-hand side variables in the two equations, (1) and (3), are identical (Wooldridge, 2002).

In our study, however, the set of regressors in each equation is different: clinical process quality in equation (1) is constructed from heart failure data, and in equation (3) from HF, AMI, and PN data; the same applies to equations (2) and (4). To check the appropriateness of SUR over OLS regression, we use the Breusch-Pagan LaGrange Multiplier test. The result indicates that the two residual terms are highly correlated ($\chi^2 = 25.64$, $p < 0.001$), supporting the use of the SUR approach. We use Huber–White standard errors and cluster them by state to control for any unobserved between-group heteroskedasticity because hospitals in different states are exposed to different state policies, health environments, and, potentially, patient demographics. We believe that multicollinearity is not a serious concern as all variance inflation factors are below 5. To summarize, we use SUR regression analysis with Huber-White standard errors clustered by state, and include state fixed effects using state dummies.

2.4 Results

A correlation matrix of the study variables (Table 2.3) shows that focus and patient experience have different associations with each of the dependent variables. While focus is positively related to readmission rates ($\rho = 0.20$, $p < 0.01$), its relationship with patient satisfaction is negative ($\rho = -0.20$, $p < 0.01$). Additionally, we observe that patient experience has a negative relationship with readmission rates ($\rho = -0.24$, $p < 0.01$), but a strong positive correlation with patient satisfaction ($\rho = 0.76$, $p < 0.01$). Lastly, our main interests, focus and patient experience, have a weak relationship ($\rho = 0.10$, $p < 0.01$).

SUR results for readmission and patient satisfaction are reported in Tables 2.4 and 2.5 respectively. For each dependent variable, we first include control variables (Model 1). We then include focus (Model 2), patient experience (Model 3), and both focus and patient experience (Model 4) in a stepwise manner to test direct effects. Finally, we evaluate the indirect effects of focus and patient experience by including the interaction (Model 5), and the absolute difference (Model 6) terms. To ensure that the joint effects are not due to non-linearity of the main effects, we incorporate quadratic terms of focus and patient experience in Model 5 (Dawson, 2014; Ganzach, 1997).

We first describe the direct effects of focus and patient experience on readmission (Table 2.4). We find that focus is positively associated with readmission ($b = 0.159$, $p < 0.001$, Model 2), resulting in a failure to support Hypothesis 1a, which theorized a negative effect. The coefficient for patient experience is negative and statistically significant ($b = -0.097$, $p < 0.001$, Model 3) suggesting a strong support for Hypothesis 2a. These effects remain negative in all other models. Our results show that both the combined and imbalance effects are significant. However, the coefficient of the interaction term is negative ($b = -0.041$, $p < 0.01$, Model 5) suggesting that our results support Hypothesis 3, which theorized a negative effect. Finally, we evaluate the imbalance effect: the coefficient for absolute difference is positive ($b = 0.064$, $p < 0.001$, Model 6), providing support for Hypothesis 5. This implies that the more divergent a hospital's attention to focus and patient experience, the worse (higher) its readmission rates.

Moving to patient satisfaction (Table 2.5), we find a significant negative association for focus ($b = -0.164$, $p < 0.001$, Model 2) and a positive association for patient experience ($b = 0.956$, $p < 0.001$, Model 3). These results provide support for Hypothesis 1b and 2b, respectively. We also observe that the interaction term is not significant ($b = 0.021$, $p = 0.225$, Model 5), indicating failed support for Hypothesis 4. In contrast, the coefficient for absolute difference term is significant ($b = -0.105$, $p < 0.001$, Model 6), providing support for Hypothesis 6. This indicates that the greater the imbalance between focus and patient experience, the worse (lower) its patient satisfaction.

Robustness Checks

We conduct several robustness checks related to endogeneity, data, measures, and other underlying assumptions, to ensure the consistency of our results. We first address endogeneity concerns by examining the relationship between focus and readmission rates. Previous literature finds that a volume-type measure of focus, such as that used in our main analysis, may be endogenous because it may reflect a selective referral effect (Huesch, 2009). Hospitals that are well-known for heart failure treatment may also attract sicker heart failure patients. These effects may result in biased estimates on focus. To address this potential bias, we use an instrumental variable two-Stage least squares (2SLS) approach (Angrist & Imbens, 1995). Because we use hospital-level data, we consider the number of cardiologists in the hospital's state as an instrument for focus. We obtain cardiologist data from *Physician Characteristics and Distribution in the US,* published by the American Medical Association. We conduct our analysis in two stages (see Appendix A, Table 3). In stage 1, we regress endogenous independent variable, focus, using the instrumental variable (number of state cardiologists), and control variables, and compute the predicted values of focus. In stage 2, we use the predicted values of focus obtained in stage 1 and the control variables to estimate the coefficients for Model 2 with SUR regression and Huber-White robust standard errors. Based on the Durbin–Wu–Hausman endogeneity test ($\chi^2 = 0.17$, $p = 0.678$), we can conclude that focus is not endogenous in the readmission rates model.

We check to ensure that our instrument satisfies relevance and exclusion restrictions assumptions (Wooldridge, 2002). Instrument relevancy requires that an instrument should be strongly correlated with the endogenous regressor in the first stage. However, the correlation between focus and number of state cardiologists is significant but relatively low ($\rho = 0.089$). Instrument exogeneity mandates that an instrument is uncorrelated with the error term in stage 2 (Wooldridge, 2002). The F-test statistic from the test of excluded instruments assumption ($R^2 = 0.346$, $F = 184.96$, $p < 0.001$) and minimum eigenvalue statistic ($F = 109.85$) are larger than the cut-off values of 10 (Staiger & Stock, 1997) and 16.38 (Stock & Yogo, 2005). These tests support the requirements of a strong instrument (Cameron & Trivedi, 2010).

Next, we check whether the results are robust to alternate measures for patient satisfaction and patient experience. First, we change the patient satisfaction measure from the overall hospital rating item (an 11-point scale) to the "willingness to recommend" item (a 4-point scale). Results using this alternative measure are substantively the same. For patient experience, we consider an alternative measure by replacing the patient experience measurement items with the items used to measure experiential quality (Chandrasekaran et al., 2012; Nair et al., 2013; Senot et al., 2016a, 2016b). The patient experience measure in this study is more encompassing and constructed using "care and non-care" aspects while the experiential quality measure used in related studies focuses solely on care (interpersonal) measures. The results with experiential quality are consistent with our reported results. It is worth noting that our measure explains 3 percent more variation in patient satisfaction than experiential quality (alternate measure).

Because our main analysis used only two periods due to three-year rolling readmission rates, we use the extended data of four time-periods to test robustness. Before conducting the analysis, we use an autocorrelation test to check for serial correlation (Wooldridge, 2002). We employ feasible generalized least squares (FGLS) to address serial correlation in the new data.[6] We report results for readmission (Models 1–4) and patient satisfaction (Models 5–8) in Table 2.6. The results with FGLS are consistent with, and marginally stronger than SUR. We find that the interaction term for predicting patient satisfaction, which was insignificant for SUR is significant now in the hypothesized direction (Model 6).

We further address between-group heteroscedasticity by repeating all our SUR regression models with Huber–White standard errors clustered by hospital. We continue to include state fixed effects using state dummies, along with all the other substantive variables. The results are identical to the main analysis.

Lastly, there may be potential bias due to common method variance (CMV) since the respondents of patient experience and patient satisfaction are the same (Chandrasekaran et al., 2012). However, we note that the effect of CMV decreases when more independent variables are added, and while CMV cannot produce any placebo effects for interaction terms, it can underestimate those effects (Siemsen et al., 2010). Therefore, the significant effects of the interaction terms even if CMV were a concern are the strong evidence that supports our hypotheses of combined effects.

[6] We remove 196 hospitals from the original sample because FGLS requires a balanced panel.

Table 2.3 Correlation Matrix for Variables

		1	2	3	4	5	6	7	8	9	10
1	*Readmission*	1.00									
2	*Satisfaction*	-0.35***	1.00								
3	*Focus*	0.20***	-0.20***	1.00							
4	*Patient Experience*	-0.24***	0.76***	0.10***	1.00						
5	*Bed* (logged)	-0.02*	-0.03**	-0.43***	-0.40***	1.00					
6	*Bed occupancy* (%)	0.08***	-0.03**	-0.34***	-0.38***	0.61***	1.00				
7	*Case Mix Index*	-0.25***	0.23***	-0.52***	-0.20***	0.67***	0.52***	1.00			
8	*Nursing Intensity Rate*	-0.12***	0.10***	-0.32***	-0.17***	0.37***	0.32***	0.39***	1.00		
9	*CPQ* (*HF*)	-0.22***	0.19***	-0.14***	0.07***	0.15***	0.07***	0.21***	0.19***	1.00	
10	*CPQ* (*Overall*)	-0.29***	0.25***	-0.16***	0.10***	0.19***	0.09***	0.28***	0.24***	0.78***	1.00

*** p<0.01, ** p<0.05 * p<0.1.

Table 2.4 Seemingly Unrelated Regression Analysis Results: Readmission Rates

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Focus		0.159***		0.153***	0.201***	0.136***
		(0.022)		(0.022)	(0.026)	(0.021)
Focus^2					-0.025***	
					(0.005)	
PatExp			-0.097**	-0.083**	-0.083***	-0.080***
			(0.030)	(0.027)	(0.019)	(0.024)
PatExp^2					0.039***	
					(0.005)	
*Focus*PatExp*					-0.041**	
					(0.015)	
|*Focus-PatExp*|						0.064***
						(0.018)
For-profit	0.271***	0.260***	0.227***	0.223***	0.201***	0.219***
	(0.038)	(0.039)	(0.039)	(0.041)	(0.040)	(0.040)
Government	0.069	0.031	0.065	0.029	0.015	0.022
	(0.043)	(0.042)	(0.041)	(0.041)	(0.037)	(0.039)
Teaching	0.064*	0.043	0.061†	0.041	0.042	0.038
	(0.032)	(0.031)	(0.031)	(0.030)	(0.030)	(0.031)
Location	0.021	0.014	-0.001	-0.004	-0.014	-0.014
	(0.043)	(0.041)	(0.040)	(0.039)	(0.038)	(0.039)
Hospital size	0.118***	0.138***	0.082*	0.107**	0.121***	0.120***
	(0.031)	(0.031)	(0.034)	(0.034)	(0.034)	(0.036)
Bed occ. rate	0.101***	0.114***	0.088***	0.103***	0.098***	0.103***
	(0.020)	(0.021)	(0.020)	(0.021)	(0.020)	(0.021)
CMI	-0.306***	-0.245***	-0.286***	-0.230***	-0.218***	-0.244***
	(0.034)	(0.031)	(0.037)	(0.033)	(0.031)	(0.034)
Nursing intensity	-0.018	-0.005	-0.017	-0.006	-0.009	-0.004
	(0.017)	(0.016)	(0.017)	(0.016)	(0.016)	(0.016)
CPQ (*HF*)	-0.022†	-0.017	-0.016	-0.012	-0.010	-0.009
	(0.013)	(0.012)	(0.013)	(0.012)	(0.011)	(0.012)
Constant	0.253***	0.395***	0.267***	0.402***	0.455***	0.329***
	(0.034)	(0.039)	(0.031)	(0.038)	(0.043)	(0.037)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,862	5,862	5,862	5,862	5,862	5,862
Chi^2	2778.2	3108.6	2868.2	3176.1	3397.0	3211.9
R^2	0.322	0.347	0.329	0.351	0.367	0.354

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Table 2.5 Seemingly Unrelated Regression Analysis Results: Patient Satisfaction

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Focus		-0.164***		-0.091***	-0.108***	-0.064**
		(0.036)		(0.023)	(0.022)	(0.022)
Focus^2					0.016***	
					(0.004)	
PatExp			0.956***	0.948***	0.983***	0.945***
			(0.051)	(0.050)	(0.020)	(0.043)
PatExp^2					-0.079***	
					(0.007)	
*Focus*PatExp*					0.021	
					(0.018)	
|*Focus-PatExp*|						-0.105***
						(0.026)
For-profit	-0.592***	-0.579***	-0.149**	-0.146***	-0.107**	-0.138***
	(0.072)	(0.072)	(0.037)	(0.038)	(0.035)	(0.036)
Government	-0.133**	-0.095*	-0.098***	-0.077***	-0.050***	-0.067***
	(0.042)	(0.040)	(0.021)	(0.019)	(0.015)	(0.007)
Teaching	-0.041	-0.020	-0.010	0.002	0.003	0.007
	(0.054)	(0.052)	(0.026)	(0.025)	(0.021)	(0.023)
Location	-0.049	-0.042	0.170***	0.172***	0.195***	0.189***
	(0.056)	(0.052)	(0.025)	(0.025)	(0.022)	(0.024)
Hospital size	-0.297***	-0.319***	0.052*	0.037†	0.023	0.015
	(0.027)	(0.029)	(0.021)	(0.022)	(0.017)	(0.023)
Bed occ. rate	0.006	-0.007	0.129***	0.121***	0.129***	0.121***
	(0.026)	(0.027)	(0.013)	(0.014)	(0.011)	(0.013)
CMI	0.367***	0.305***	0.173***	0.140***	0.134***	0.164***
	(0.024)	(0.023)	(0.018)	(0.014)	(0.014)	(0.015)
Nursing intensity	0.046*	0.033†	0.046**	0.039**	0.041**	0.038*
	(0.020)	(0.020)	(0.016)	(0.015)	(0.015)	(0.015)
CPQ (*All*)	0.151***	0.141***	0.061***	0.056***	0.042***	0.051***
	(0.022)	(0.020)	(0.012)	(0.011)	(0.010)	(0.010)
Constant	0.063	-0.086†	-0.107***	-0.188***	-0.211***	-0.070†
	(0.048)	(0.047)	(0.025)	(0.028)	(0.034)	(0.037)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,862	5,862	5,862	5,862	5,862	5,862
Chi^2	1684.9	1873.9	19160.3	19754.2	21006.9	19755.0
R^2	0.223	0.242	0.766	0.771	0.782	0.771

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Table 2.6 Robustness Checks

	Readmission Rates				Patient Satisfaction			
	Combined		Imbalance		Combined		Imbalance	
	SUR	FGLS	SUR	FGLS	SUR	FGLS	SUR	FGLS
	1	2	3	4	5	6	7	8
Focus	0.201***	0.256***	0.136***	0.187***	-0.108***	-0.123***	-0.064**	-0.079***
	(0.026)	(0.005)	(0.021)	(0.005)	(0.022)	(0.002)	(0.022)	(0.002)
$Focus^2$	-0.025***	-0.030***			0.016***	0.016***		
	(0.005)	(0.002)			(0.004)	(0.001)		
PatExp	-0.083***	-0.066***	-0.080***	-0.067***	0.983***	0.873***	0.945***	0.866***
	(0.019)	(0.004)	(0.024)	(0.005)	(0.020)	(0.002)	(0.043)	(0.002)
$PatExp^2$	0.039***	0.038***			-0.079***	-0.065***		
	(0.005)	(0.002)			(0.007)	(0.001)		
*Focus*PatExp*	-0.041**	-0.033***			0.021	0.017***		
	(0.015)	(0.004)			(0.018)	(0.001)		
|*Focus-PatExp*|			0.064***	0.052***			-0.105***	-0.083***
			(0.018)	(0.004)			(0.026)	(0.002)
For-profit	0.201***	0.217***	0.219***	0.231***	-0.107**	-0.152***	-0.138***	-0.170***
	(0.040)	(0.009)	(0.040)	(0.011)	(0.035)	(0.004)	(0.036)	(0.004)
Government	0.015	-0.034***	0.022	-0.037***	-0.050***	-0.059***	-0.067***	-0.063***
	(0.037)	(0.010)	(0.039)	(0.010)	(0.015)	(0.004)	(0.007)	(0.004)
Teaching	0.042	0.102***	0.038	0.101***	0.003	-0.024***	0.007	-0.021***
	(0.030)	(0.008)	(0.031)	(0.009)	(0.021)	(0.003)	(0.023)	(0.003)
Location	-0.014	-0.052***	-0.014	-0.040***	0.195***	0.190***	0.189***	0.181***
	(0.038)	(0.010)	(0.039)	(0.011)	(0.022)	(0.004)	(0.024)	(0.004)
Hospital size	0.121***	0.194***	0.120***	0.198***	0.023	-0.020***	0.015	-0.029***
	(0.034)	(0.006)	(0.036)	(0.006)	(0.017)	(0.002)	(0.023)	(0.002)
Bed occ. rate	0.098***	0.145***	0.103***	0.158***	0.129***	0.099***	0.121***	0.090***
	(0.020)	(0.004)	(0.021)	(0.005)	(0.011)	(0.002)	(0.013)	(0.002)
CMI	-0.218***	-0.301***	-0.244***	-0.335***	0.134***	0.193***	0.164***	0.226***
	(0.031)	(0.006)	(0.034)	(0.007)	(0.014)	(0.002)	(0.015)	(0.002)
Nursing intensity	-0.009	-0.025***	-0.004	-0.021***	0.041**	0.051***	0.038*	0.050***
	(0.016)	(0.005)	(0.016)	(0.005)	(0.015)	(0.002)	(0.015)	(0.002)
CPQ (*HF*)	-0.010	-0.008*	-0.009	-0.011**	0.042***	0.036***	0.051***	0.042***
	(0.011)	(0.003)	(0.012)	(0.003)	(0.010)	(0.001)	(0.010)	(0.001)
Constant	0.455***	0.365***	0.329***	0.307***	-0.211***	-0.035***	-0.070†	0.008*
	(0.043)	(0.009)	(0.037)	(0.011)	(0.034)	(0.004)	(0.037)	(0.004)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of hospitals	3,027	2,831	3,027	2,831	3,027	2,831	3,027	2,831
Observations	5,862	11324	5,862	11,324	5,862	11,324	5,862	11,324
Wald-Chi^2 or (*F*)	3397.0	146958.0	3211.9	71647.7	1684.9	238330.7	19755.0	228380.1
R^2	0.367	–	0.354	–	0.223	–	0.771	–

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.
Notes. We consider AR(1) in FGLS and robust standard errors in parentheses.

2.5 Discussion

We summarize our results in Table 2.7. Overall, we show that focus increases readmission rates and reduces patient satisfaction, both of which are undesirable outcomes from managerial, regulatory, and patient perspectives. In contrast, patient experience has a desirable effect on both outcomes, by resulting in lower readmission rates and increased patient satisfaction. We also find that focus and patient experience together reduce readmission rates, implying that patient experience mitigates the negative effect of focus on readmission rates. Finally, an imbalance between focus and patient experience results in an increase in readmission rates and a reduction in patient satisfaction; both are undesirable outcomes. To further investigate these effects, we conduct detailed post-hoc analysis. We describe the theoretical, managerial, and policy implications of these results below.

Table 2.7 Summary of Results

	Readmission Rates		**Patient Satisfaction**	
	Direction*	Desirable	Direction*	Desirable
Focus	Positive	No	Negative	No
Patient Experience	Negative	Yes	Positive	Yes
*Focus*Patient Experience*	Negative	Yes	–	–
|*Focus – Patient Experience*|	Positive	No	Negative	No

*Only statistically significant effects from Tables 2.4 and 2.5 are shown.

Our research makes three notable contributions to existing theory. First, by showing that focus has an undesirable effect on both readmission rates and patient satisfaction, our results reveal that focus as a strategy has an undetected dark side which previous research has failed to identify in the healthcare setting. Instead, previous studies have predominantly highlighted positive effects of focus on efficiency (e.g. cost) and clinical performance (e.g. mortality rate) (Clark & Huckman, 2012; KC & Terwiesch, 2011; Lee et al., 2015). Until the current study, few studies have examined the relationship between focus and readmission rates or overall patient satisfaction for the entire hospital. We explore the boundary conditions where these results apply, and also identify an underlying mechanism that might explain the undesirable relationship between focus and readmission rates.

To better understand the boundary conditions, we examine whether the direct effects of focus vary across different types of hospitals, based on two hospital characteristics: resource utilization and teaching status. Researchers have studied these two characteristics extensively and find they impact hospital performance systematically and significantly (Goldstein & Iossifova, 2012; McDermott & Stock, 2011). We measure hospital resource utilization by average bed occupancy rates, and categorize our sample hospitals into two groups: hospitals with high and low bed occupancy. Teaching status is measured as a binary variable (Yes/No), and hospitals are divided into two groups: teaching and non-teaching hospitals. For each group, we replicate Models 2−4 of our main SUR analysis.

For readmission rates (see Appendix A, Table 4), our results show that the direct effect of focus is positive and significant for hospitals with both 'high' bed occupancy ($b = 0.228$, $p < 0.001$) and 'low' bed occupancy ($b = 0.090$, $p < 0.001$). A beta coefficient difference test shows that the effect is significantly greater in the high occupancy group ($p < 0.001$). This makes sense because higher utilization requires hospitals to consume substantial resources for inpatients in the focal area. We conducted the exact same analysis for teaching status and find similar results (see Appendix A, Table 5). The direct effect of focus on readmission rates is positive and significant for teaching hospitals ($b = 0.222$, $p < 0.001$) and non-teaching hospitals ($b = 0.123$, $p < 0.001$), and the difference between them is also significant ($p = 0.018$).

In regards to patient satisfaction, our main results showed that hospitals with greater focus have lower patient satisfaction (Table 2.5). To understand the generalizability of this result, we repeat the analysis using the two hospital groups categorized by bed occupancy and teaching status (see Appendix A, Table 6). These results show that the direct effect of focus is significant and negative for both occupancy groups (high: $b = -0.224$, $p < 0.001$; low: $b = -0.128$, $p < 0.001$), and the effect is again greater for the high occupancy group ($p = 0.001$). For teaching status (see Appendix A, Table 7), our results show the focus has a negative effect in both groups (teaching: $b = -0.309$, $p < 0.001$; non-teaching: $b = -0.114$, $p < 0.001$), and the effect is greater for teaching hospitals ($p = 0.002$).

We also conducted the subgroup analyses to examine the boundary conditions associated with the positive effect of patient experience on readmission rates and patient satisfaction. The results show that the effects are significant in hospitals with both high and low occupancy rates, as well as teaching and non-teaching hospitals, for the two performance measures. The differences are not significant across the groups, implying that patient experience has a uniformly desirable effect on both measures under the tested boundary conditions.

Lastly, to understand why analysis shows a positive relationship between focus and readmission, we explore a possible mechanism. Our literature review revealed that, while there is limited research linking focus with readmission rates, prior researchers have shown that focus reduces mortality rates (Ding, 2015; Lee et al., 2015), and in separate studies, other researchers have shown mortality reduces readmission rates (Jha, 2018; Press et al., 2013). While it is logical that high mortality results in lower readmission rates, the low mortality-high readmission relationship may require an explanation. Jha (2018) conjectured that hospitals with low mortality rates might be better at keeping their sickest patients alive, who are at higher risk of being readmitted than the average patient, which increases readmission rates. They concluded that for conditions such as heart failure, it should not be surprising that hospitals with low mortality rates have higher readmission rates. In light of these observations, we propose that mortality rates might explain the positive relationship between focus and readmission by acting as a mediator between them. We test our supposition using a mediation test.

Three things are generally required to demonstrate mediation: (1) the explanatory variable that predicts the mediator; (2) the mediator predicting the dependent variable; and (3) the decrease in the direct effect of the explanatory variable when the model incorporates the mediator with the explanatory variable (Baron & Kenny, 1986). Following this sequence requires conducting three separate regression analyses. We obtain mortality rate for the hospitals in our sample, and use SUR to run the analysis (Table 2.8). We first use mortality rate as the dependent variable, and find that focus has a significant negative effect ($b = -0.184$, $p < 0.001$, column 1). Next, we use focus to predict readmission rates ($b = 0.161$, $p < 0.001$, column 2). Lastly, we incorporate the mortality rate into the full SUR model and find that it is negatively associated with readmission rates ($b = -0.072$, $p < 0.01$, column 3). We also observe that the beta coefficient for focus is marginally smaller than in column 1 but it retains significance at 0.001 level.

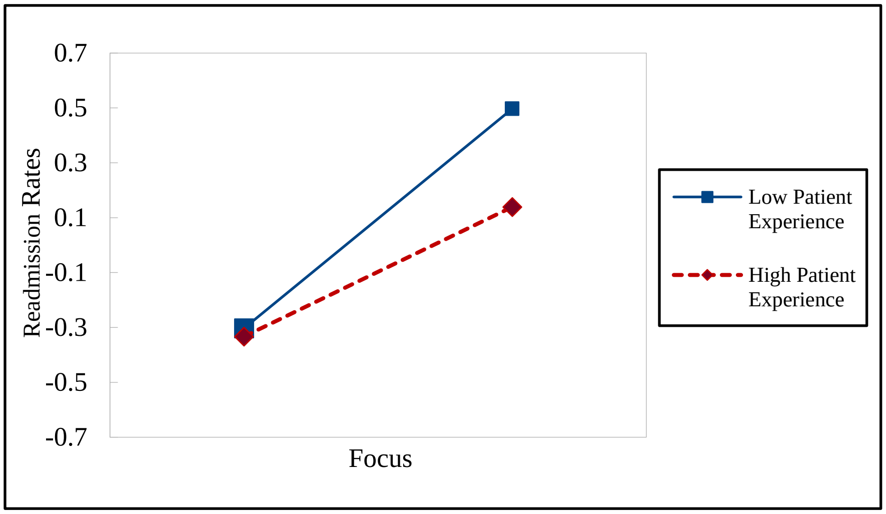
We conduct additional subgroup analysis to test whether the mediation effect differs for high and low mortality hospitals (statistical results not included here). We find that the mediation effect is consistent across both high and low mortality hospitals, and that the effect size does not differ across these groups. Taken together, this provides strong evidence that mortality rate mediates the relationship between focus and readmission rates.

Table 2.8 Mediation Test Using Mortality Rate as the Mediator

	Column 1 Mortality	Column 2 Readmission	Column 3 Readmission
Focus	-0.184***	0.161***	0.147***
	(0.020)	(0.022)	(0.020)
Mortality			-0.072***
			(0.015)
For-profit	-0.153***	0.255***	0.244***
	(0.039)	(0.039)	(0.038)
Government	0.060†	0.035	0.040
	(0.036)	(0.042)	(0.041)
Teaching	-0.053†	0.057†	0.053†
	(0.028)	(0.030)	(0.030)
Location	-0.162***	0.021	0.010
	(0.045)	(0.042)	(0.041)
Hospital size	-0.060†	0.112***	0.108***
	(0.032)	(0.027)	(0.027)
Bed occ. rate	-0.076**	0.121***	0.116***
	(0.029)	(0.022)	(0.020)
CMI	-0.077**	-0.229***	-0.235***
	(0.023)	(0.029)	(0.028)
Nursing intensity	0.015	0.006	0.007
	(0.013)	(0.017)	(0.017)
CPQ (*HF*)	-0.008	-0.017	-0.018
	(0.013)	(0.012)	(0.012)
Constant	0.533***	0.379***	0.417***
	(0.036)	(0.039)	(0.041)
Time FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Observations	5,848	5,848	5,848
F	21.85	65.15	65.21
R^2	0.185	0.403	0.407

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Figure 2.1 Moderation Plot: Focus and Patient Experience for Readmission Rates



Our second contribution consists in showing that, together, focus and patient experience reduce readmission rates (indicated by negative coefficient in Model 5, Table 2.4). Considering the positive effect of focus and the negative effect of patient experience, this joint effect suggest that patient experience mitigates the negative influence of focus on readmission. To better understand the nature of the joint effects on readmission, we examine the interaction plot (Figure 2.1). We find that as focus increases, readmission rates increase regardless of the level of patient experience, as indicated by the upward sloping solid and dashed lines representing low and high levels of patient experience, respectively in Figure 2.1. Further, we observe that a low level of patient experience (i.e. the solid line) has a steeper slope and a higher end-point of readmission compared to a high level of patient experience (dashed line). This pattern of results suggests that together, focus with low level of patient experience is particularly harmful to readmission rates.

Although non-linear effects were not hypothesized, focus and patient experience quadratic terms are included to ensure the accuracy of the beta coefficients associated with the interaction term in equation 1. Both quadratic terms ($Focus^2$ and $PatExp^2$) are statistically significant but in opposite directions (Model 4, Table 2.4). $Focus^2$ has a negative sign, and along with its positive linear effect, indicates a negative U-shaped relationship between focus and readmission. This implies that low and high levels of focus are associated with low readmission rates, while moderate levels of focus result in relatively higher readmission rates. In contrast, $PatExp^2$ is positive, while its direct effect is negative. This suggests a U-shaped relationship between patient experience and readmission, where low and high levels of patient experience are less desirable than moderate levels. Taken together, the non-linear and joint effects demonstrate the nuanced nature of relationships between focus and patient experience, and their impact on readmission rates. While our results paint a complicated picture, they are important for all stakeholders because decisions made using results from prior studies may be incomplete and inaccurate as none of them include both focus and patient experience, or examine their joint effects, in one single study. Our study is the first step toward linking previously unstudied initiatives and their impacts.

Third, our results related to the imbalance effect show that keeping balance between focus and patient experience is also important because favoring either focus or patient experience over the other degrades both types of performance, by increasing readmission and reducing patient satisfaction. Recall that imbalance, measured as the absolute difference between focus and patient experience, does not distinguish between positive and negative differences. We use a spline regression model to identify whether the negative influence is exacerbated when one of these is emphasized over the other. Spline regressions are frequently used to establish a point (or multiple points) where a continuous relationship changes slope at a 'knot' (Marsh & Cormier, 2002). A major advantage of spline regression is that it does not reduce sample size and attenuate a continuous variable. We adopt a spline function with a single knot and use one dummy variable to signify values above and below the knot. We define the difference between focus and patient experience as $DF_{it} = Focus_{it} - PatExp_{it}$, and compute an indicator variable (D) which takes the value of zero when the difference (DF_{it}) is negative and one otherwise. Specifically, DF is negative when hospitals emphasize patient experience over focus and positive when hospitals put more emphasis on focus than patient experience. The following set of equations represent these conditions:

$$Readmission\ Rate_{it} = \beta_0 + \beta_1 Focus_{it} + \beta_2 PatExp_{it} + \beta_{31}|DF_{it}| + \beta_{32}D(DF_{it} - 0) + \beta_4 C_{hf,it} + \boldsymbol{X}_{it}\gamma + \varepsilon_{it} \quad (5)$$

$$where\ DF_{it} = Focus_{it} - PatExp_{it}$$

(i) *For* $DF_{it} < 0$ ($D = 0$),

$$Readmission\ Rate_{it} = \beta_0 + \beta_1 Focus_{it} + \beta_2 PatExp_{it} - \beta_{31} DF_{it} + \beta_4 C_{hf,it} + \boldsymbol{X}_{it}\gamma + \varepsilon_{it} \quad (5a)$$

(ii) *For* $DF_{it} \geq 0$ ($D = 1$),

$$Readmission\ Rate_{it} = \beta_0 + \beta_1 Focus_{it} + \beta_2 PatExp_{it} + (\beta_{31} + \beta_{32})\, DF_{it} + \beta_4 C_{hf,it} + \boldsymbol{X}_{it}\gamma + \varepsilon_{it} \quad (5b)$$

We conduct a similar analysis for patient satisfaction. The results of both of these models are presented in Table 2.9. First, we note that the imbalance effects on readmission rates (b_{31} = 0.066, $p < 0.001$) and on patient satisfaction ($b_{31} = -0.108$, $p < 0.001$) even after including D, the indicator spline variable, are consistent with our main models, both in algebraic sign and statistical significance. Next, we note that the indicator variable is marginally significant in both the readmission rates and patient satisfaction models. Specifically, we find that the indicator variable in the readmission model is positive (b_{32} = 0.063, p = 0.089), which indicates that favoring focus over patient experience is associated with a higher readmission rates. In contrast, the negative sign in the patient satisfaction model (b_{32} = −0.084, p = 0.060) shows that favoring focus over patient satisfaction results in lower patient satisfaction. We used balance plots to further substantiate these results. In Figure 2.2, we show that an imbalance created by high focus increases readmission rates substantially (i.e. moving left to right on the x-axis), while imbalance from high patient experience does not increase readmission rates (moving bottom to top on the y-axis). If an increase in focus is accompanied with an increase in patient experience (i.e. imbalance is reduced), readmission rates remain the same. A similar pattern is observed in relation to patient satisfaction (Figure 2.3). However, patient satisfaction increases slightly with a reduction in imbalance.

Figure 2.2 Balance Plot: Focus and Patient Experience for Readmission Rates

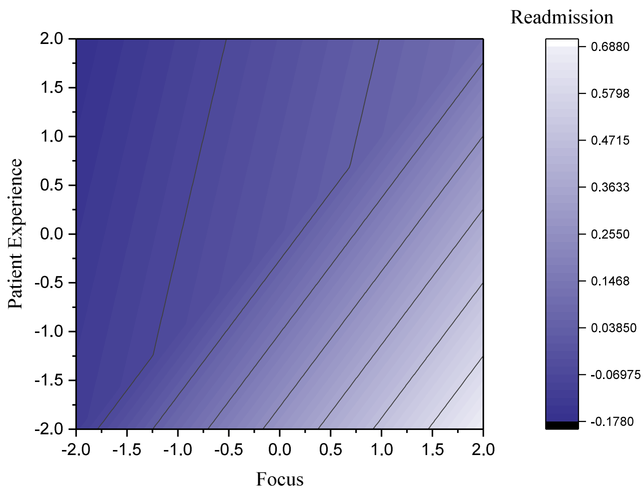


Figure 2.3 Balance Plot: Focus and Patient Experience for Satisfaction

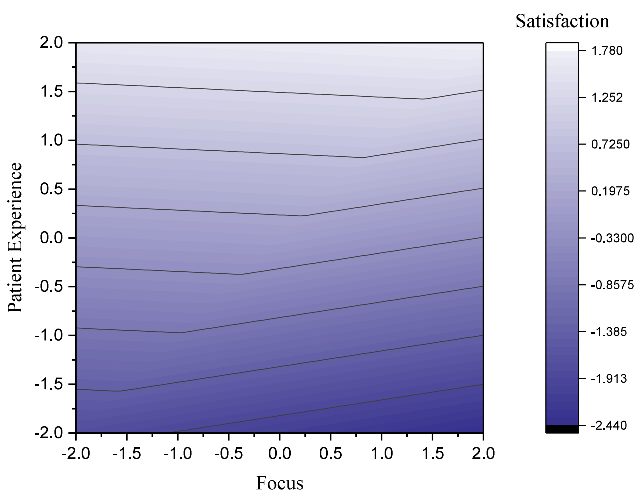


Table 2.9 Spline Regression using Seemingly Unrelated Regression

	Readmission	Satisfaction
	Model 1	Model 2
Focus	0.119***	-0.041
	(0.023)	(0.028)
PatExp	-0.063*	0.921***
	(0.025)	(0.050)
|*Focus-PatExp*| (β_{31})	0.066***	-0.108***
	(0.017)	(0.026)
Spline Dummy (β_{32})[a]	0.063†	-0.084†
	(0.037)	(0.045)
For-profit	0.216***	-0.135***
	(0.040)	(0.035)
Government	0.022	-0.066***
	(0.039)	(0.017)
Teaching	0.038	0.007
	(0.031)	(0.023)
Location	-0.016	0.191***
	(0.039)	(0.024)
Hospital size	0.123***	0.012
	(0.036)	(0.023)
Bed occ. rate	0.104***	0.120***
	(0.021)	(0.013)
CMI	-0.245***	0.165***
	(0.034)	(0.015)
Nursing intensity	-0.004	0.038*
	(0.016)	(0.015)
CPQ (*HF*)	-0.009	
	(0.012)	
CPQ (*All*)		0.051***
		(0.010)
Constant	0.306***	-0.040
	(0.040)	(0.046)
Time FE	Yes	Yes
State FE	Yes	Yes
Observations	5,862	5,862
F	64.23	377.98
R^2	0.411	0.804

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

a: The positive sign in the readmission rates model indicates that favoring focus over patient experience is associated with higher readmission rates (b_{32} = 0.063, p = 0.089). The negative sign in the patient satisfaction model shows that favoring focus over patient experience results in lower patient satisfaction (b_{32} = −0.084, p = 0.060).

These results both validate our main results, and also help to isolate where and when the relationships occur. As a set, our results show that while any imbalance appear undesirable, it is in fact hospitals which favor focus over patient experience that suffer greater degradation in their performance, compared to the inverse where hospitals favor patient experience over focus.

These results have critical implications for hospital management and regulators. Our study shows that managers face challenging tradeoffs in their pursuit to improve multiple dimensions of performance. A focus strategy has been associated with superior performance in numerous empirical studies, and has resulted in an inordinate emphasis on a few selected clinical specialties in many acute care hospitals. However, our study shows that focus has a significant dark side that was not recognized in previous studies. We attribute this to a predominance of efficiency related performance measures such as cost and length of stay in prior studies. This negative effect of focus may be more salient in acute care general hospital settings, where management is required to offer a broad set of services.

The dark side of focus is buttressed by the spline regression results, which show that emphasizing focus over patient experience has an undesirable impact on both readmission and patient satisfaction. As a set, our results suggest that management must critically evaluate how much to emphasize focus, especially if the objective is to improve non-efficiency oriented performance measures. In contrast, by substantiating the implicit influence of patient experience, we not only provide evidence of its broader benefits but also show that it may help to mitigate the negative impact of focus on both types of performance measures. At a higher level, our results imply that investing resources in practices that improve care and non-care aspects of patient experience have positive spillover effects beyond patient satisfaction.

Finally, the results have significant implications for regulators responsible for designing policies to encourage hospital administrators to simultaneously improve multiple performance measures. Our results show that it may be very challenging to do so because no single mechanism positively impacts multiple performance measures. Our results also suggest that performance evaluation policies should be contingent on the type of hospitals. Specifically, specialty hospitals face less complex challenges and should not be evaluated in the same manner as a general acute care hospital. Recognizing these challenges should help regulators design policies that account for the inherent tradeoffs among different performance measures on the one hand, and accommodate different types of hospitals on the other.

2.6 Conclusion and Limitations

Although focus and patient experience in hospitals have attracted immense attention from researchers, hospital administrators, and policy makers, our current understanding lacks a coherent theory to clearly explain their relationship with hospital performance. We address this research gap by providing a holistic view of these relationships. Our results underscore not only the dark side of focus, but also the beneficial effects of patient experience on readmission rates and patient satisfaction. Even so, our study has some limitations related to data and measures that might offer opportunity for future research.

First, we use aggregated hospital level data. Some previous studies have used patient level data (KC & Terwiesch, 2011) or patient level data nested in hospitals (McDermott & Stock, 2011). However, readmission rates and overall patient satisfaction are aggregated from patient level responses, and should not significantly impact the relationships. Second, while our data span 2007 to 2013, we only have two time windows because HCAHPS provides rolling data over a three-year window. Although, we validate our results with an extended four-period model, it would be useful to have longer panel or annual data. Our data are from 3027 acute care facilities, without specialty free-standing hospitals, which would be useful for understanding the generalizability of our results. Finally, for our unit of analysis, we measure focus and readmission for "heart-failure (HF)" patients, while patient experience and patient satisfaction are measured for all inpatients. While this is logical for hypotheses 1a and 2b, which link variables in the same sets, it may pose a challenge when evaluating hypotheses 1b and 2a, which link focus with patient satisfaction, and patient experience with readmission, respectively. In hypothesis 1b, it is likely that the effect is underestimated due to the smaller HF patient set; but the strong significance of our results should ameliorate the "unit of analysis" concern. In case of hypothesis 2b, HF patients constitute a part of the overall inpatient population, and clinical and administrative staff are trained to interact with all patients in the same manner. Again, the differing patient sets used to measure patient experience and readmission should result in underestimation of the effect. This conservative approach has an advantage: using different units of analysis, and hence different respondents, reduces the potential for common method bias.

We expect our study to spur other researchers to more thoroughly investigate the effect of focus and patient experience in different contexts and for different types of performance measures. For instance, one potential question is to examine whether the effect of focus changed before and after the implementation of the Affordable Care Act (ACA). Our research primarily addresses the period before the ACA. However, the Hospital Readmission Reduction Program in the ACA, implemented in 2013, shifted hospital incentives toward reducing readmission rates for evaluating the hospital for pay for performance program, making it an interesting question to study.

Notwithstanding the limitations, our results highlight the challenges that hospital administrators face in pursuing seemingly mutually exclusive performance objectives. While there is no one magic bullet to improve both readmission rates and overall patient satisfaction at the same time, our results show that patient experience has desirable direct and indirect effects because of its ability to overcome the negative effects of focus.

Chapter 3

Does Announcing the Visit Matter? An Empirical Examination in US Nursing Homes

3.1 Introduction

Various types of inspections have been used to improve and monitor quality and safety in the manufacturing and service industries (Anand et al., 2012; Ball et al., 2017; Gray et al., 2011; Levine et al., 2012; Mani & Muthulingam, 2019; Short et al., 2016). Inspecting agencies make an important choice between two inspection modes: either announce the inspection before arriving at the facility or make an unannounced inspection with little advance notification (Greenfield et al., 2012; Klerks et al., 2013). In practice, agencies use both types of strategies, and may sometimes combine the two inspection modes (GAO 2004, 2006, 2007a, 2007b, 2008, 2012a, 2012b). Table 3.1 lists the use of announced and unannounced inspections strategies at various regulatory and accrediting agencies. Some agencies, such as the Federal Highway Administration, only use the announced inspection strategy. They believe that unannounced inspections create a "gotcha" environment that damages their relationship with the inspected organizations (GAO, 2012a). However, Table 3.1 also shows that the healthcare industry often uses the unannounced inspection strategy. According to Centers for Medicare and Medicaid Services (CMS) and The Joint Commission (TJC), unannounced inspections better capture an accurate picture of the organization’s actual day-to-day processes, while announced inspections allow for temporal adjustments in organization’s daily processes (GAO, 2004, 2006, 2007b). Finally, Table 3.1 shows that some agencies, like the U.S. Coast Guard and the Department of the Interior, use both types of inspection strategies (GAO, 2008). While these agencies argue for the efficacy of unannounced inspections, they also concede that unannounced inspections create other issues. For instance, unannounced visits may occur when needed records and personnel were absent during inspections (Allen, 2015; GAO, 2007a, 2012b). In contrast, announced inspections enable organizations to assemble appropriate records and personnel prior to the visit (GAO 2012a, 2012b).

Table 3.1 Current Usage of Announced and Unannounced Inspections

Inspecting body	Announced	Unannounced	Target unit
Federal Highway Administration (FHWA)[1]	X		Highway infrastructure
Department of Education (DOE)[2]	X		Schools
Commission on Accreditation of Rehabilitation Facilities (CARF)	X		Healthcare organizations
Center for Medicare & Medicaid Services (CMS)[3]		X	Healthcare organizations
Healthcare Facilities Accreditation Program		X	Healthcare organizations
The Joint Commission (TJC)[4]	Conditional	X	Healthcare organizations
Department of Defense (DOD)[5]		X	Child development facilities
U.S. Department of Agriculture (USDA)[6]		X	Animals and plant dealers
Food and Drug Administration (FDA)[7]	Foreign region	Domestic region	Drug manufacturing plants
U.S. Coast Guard (USCG)[8]	X	X	Maritime facilities
Department of the Interior (DOI)[9]	X	X	Oil and gas facilities

US. Government Accountability Office (GAO) Reports: [1] GAO-12-474, [2] GAO/HEHS-96-143, [3] GAO-04-850 & GAO-06-416, [4] GAO-07-79, [5] GAO/HEHS-00-7, [6] GAO-10-945, [7] GAO-08-224T, [8] GAO-08-12, [9] GAO-12-423.

Despite the debates on the efficacy of these different inspection strategies, scholars have not examined the operational differences between announced and unannounced inspections. A few studies that have explored announced and unannounced inspections did not consider the impact on operational performance outcomes (e.g., quality, safety) (Ehlers et al., 2017; Greenfield et al., 2012; Klerks et al., 2013), which is the main goal in adhering to standards. In addition, an implicit assumption of unannounced inspections is that it will lead to more attention to the standard and will better sustain performance, but no coherent theory explains how announced and unannounced inspections lead to operational performance outcomes. Further, scholars have not investigated the differences between the immediate and sustained effects of an announced versus and unannounced inspection strategy on performance. Therefore, research is needed to better understand the difference between announced and unannounced inspections, and its immediate and sustained operational performance effects. This study addresses these research gaps and examines the following questions: *Do announced and unannounced inspections lead to an immediate and/or a sustained increase in overall quality performance?* To the best of our knowledge, it would be the first study to compare the operational performance differences between different inspection strategies and their short and long term operational effects.

The attention-based view (Ocasio, 1997; Shepherd et al., 2017) provides a theoretical lens to investigate this question. This theory implies that what firms do depends on what they focus their attention on (Ocasio, 1997). Drawing on this theory, we argue that announced inspection results in *transient attention* to the standard while unannounced inspection results in *sustained attention*. We operationalize operational performance as quality performance, and empirically examine immediate and sustained effects of announced and unannounced inspections on quality performance in nursing homes that are accredited by The Joint Commission (TJC). An econometric analysis of panel data from the CMS, TJC, and Long-Term Care Focus datasets spanning a 4-year period shows that both announced and unannounced inspections increase nursing home quality. However, an unannounced inspection leads to more sustained quality performance, while quality performance tends to decline after an announced inspection.

This study makes several theoretical contributions. First, it is the first research to investigate the effect of announced and unannounced inspections on operational performance. Previous literature does not go beyond examining the effect of announced and unannounced inspections to compliance of the standard. Second, it contributes to developing a theory that explains the difference between announced and unannounced inspections on performance outcomes. We provide strong theoretical explanation of how the types of inspections related to the attention in attention-based view theory and the actual practice in nursing homes. Third, this research provides empirical evidence that announced and unannounced inspections play different roles in affecting immediate and sustained quality performance.

This study also provides important managerial and practical implications. First, the results have broad applicability to inspection agencies, who want to improve and maintain quality through inspection and compliance. The analyses suggest that announced inspections result in immediate improvements in quality performance, but this is effect is more short term. Organizations can make temporal adjustments to processes prior to the inspection. In contrast, unannounced inspections tend to result in more sustained improved quality performance over time. Therefore, in the context of healthcare, unannounced inspections can be effective where sustained high-quality performance is critically important. Next, we empirically show the quality performance benefits for organizations interested in going through accreditation. The results suggest that inspections help organizations improve and sustain quality performance.

The rest of the paper has the following organization. The Sections 3.2 and 3.3 describe research context, relevant literature and hypothesis development. Section 3.4 gives the research design, empirical setting, data, and describes the empirical strategies employed. Section 3.5 gives the results and 3.6 discusses the implications. Finally, Section 3.7 discusses limitations and topics for future research.

3.2 Research Context

Healthcare organizations that have TJC accreditation provide an attractive context to study inspection strategies for several reasons. First, TJC is the largest accrediting agency that accredits more than 21,000 healthcare organizations and programs across the United States.[7] Second, inspection plays a critical role in the accreditation process at TJC.[8] Before the inspection, the organization prepares for the site visit by evaluating if they met the accreditation standards and takes appropriate actions if they have any deviations from the standard. Then, inspectors visit the organization and carry out an on-site inspection. During the inspection, inspectors interview the top management team, staff, residents/patients and their family members, trace the delivered care and treatment, and review multiple documents to ensure compliance to the criteria. Depending on the inspection outcome, TJC may require the organization to make improvements for any unmet

[7] https://www.jointcommission.org/facts_about_the_joint_commission/
[8] Surveyors (inspectors) visit healthcare organizations to evaluate compliance with pertinent standards. This visit implies an inspection and is called as a survey (http://www.jointcommission.org/about/jointcommissionfaqs.aspx).

criteria.[9] Thus, the entire accreditation process has the following phases: preparing for an inspection, receiving an inspection, and addressing non-compliance issues identified from the inspection. Third, TJC announces the first inspection for an initial application in advance, but recurring inspections are unannounced (GAO, 2004, 2007b). Therefore, TJC accredited organizations provide an attractive context to investigate announced and unannounced inspections.

However, TJC did not always conduct recurring inspections on an unannounced basis. Before 2006, all TJC inspections were announced in advance. TJC shifted from announced to unannounced regime in 2006 to follow federal government guidelines, which indicated that TJC had not identified some serious operational deficiencies that could have been identified with unannounced inspections (GAO, 2004, 2007b). TJC conceded that when healthcare organizations no longer know when inspectors will visit, they would more likely maintain compliance on an ongoing basis. TJC further postulated that an unannounced inspection will eliminate the ramp-up process and reduce unnecessary costs, which occurs when the organization view the inspection as a short-term event. Thus, TJC perceived that adopting an unannounced inspections strategy will increase the credibility of the inspection and accreditation process by disclosing a clearer picture of the actual healthcare delivery systems and actual care delivered (JCAHO & JCR, 2006). However, there was a lot of anxiety when TJC initially started conducting unannounced inspections. Some felt anxious since not having time to prepare for the inspection and not having key staff present during the inspection would lead to more deficiencies (JCAHO & JCR, 2007). Others tried to leverage this opportunity and through education empowered their staff to be more engaged in the accreditation process, which would result in more ongoing compliance (JCAHO & JCR, 2006). Overall, nobody knew exactly how this change would work.

We note that announcement timing is the main difference between announced and unannounced inspections. An announced inspection is typically scheduled one month before the actual site visit. In contrast, for an unannounced inspection, TJC notifies the relevant staff member of the inspection schedule on the morning of the inspection day, by 7:30 am in the organization's local time zone. We take this setting to examine announced and unannounced inspections on quality performance in US nursing homes.

[9] https://www.jointcommission.org/assets/1/6/2009_LTC_Overview_Combo_10_30_09.pdf

3.3 Literature Review and Hypotheses Development

Inspection Literature Review

A number of industries ranging from manufacturing, pharmaceutical, healthcare, oil and gas, to many others use inspection as an effective tool to ensure compliance (Anand et al., 2012; Ball et al., 2017; Gray et al., 2011; Levine et al., 2012; Mani & Muthulingam, 2019; Short et al., 2016). Researchers have investigated inspections in the context of occupational safety (Levine et al., 2012; Short et al., 2016), food safety (Ibanez & Toffel, 2019; Reske et al., 2007), healthcare compliance (Ehlers et al., 2017; Lu & Wedig, 2012), environmental performance (Dhanorkar et al., 2018; Mani & Muthulingam, 2019), product recall (Ball et al., 2017), and quality risk (Gray et al., 2011). Two major streams of the literature on inspections have emerged, reflecting the perspective of the inspectors and the organizations. Table 3.2 summarizes the literature from these two streams of research. The first stream of literature examines factors that influence inspectors' inspection activity and stringency. The findings indicate that market competition, work environment (e.g., weather), inspector's experience, inspector's schedule, inspector team composition, and other inspector characteristics influence inspection outcome and leniency (Ball et al., 2017; Bennett et al., 2013; Ibanez & Toffel, 2019; Scott, 2018; Short et al., 2016). The second research stream examines the relationship between inspections and organizational factors on performance outcomes. These studies show the relationship between cultural distance and inspection outcome (Gray et al., 2011), chain (network affiliation) and inspection outcome (Lu & Wedig, 2012), inspection and occupational safety (Levine et al., 2012), and R&D colocation and inspection outcome (Gray et al., 2015). A few recent studies focus on the behavioral perspective of inspected organizations. For example, Mani and Muthulingam (2018) examine the role of organizational learning (e.g., cumulative volume of inspections) on environmental performance. Specifically, they show that inspected organizations learn from direct and vicarious inspections. Dhanorkar et al. (2018) also showed that the timing of an inspection influences managerial attention in project implementation. Anand et al. (2012) studied inspections as renewal events that help prevent system deterioration. They concede that, while inspections will increase managerial attention to operational activities in the organization, but the organization will inevitably shift its attention to other priorities after the inspection.

Overall, the literature review highlights several important gaps. First, most authors have focused on the inspection compliance (e.g., number of non-compliance issues or deficiencies), but only a few have focused on operational outcomes. For instance, Ball et al (2017) examined the effect of inspections on product recalls and Levine et al., (2012) considered the effect on occupational safety (e.g., injury rate). Second, little research has investigated the types of inspections: some researchers have considered types of inspections as control variables but not as the main objective of the study. Moreover, in many studies that were included in the review, announced and unannounced inspections are poorly delineated. Instead, the authors tend to distinguish inspection types based on routine and non-routine inspections (Ibanez & Toffel, 2019; Short et al., 2016), related and non-related inspections (Dhanorkar et al., 2018), inspection outcome severity (Ball et al., 2017; Dhanorkar et al., 2018; Mani & Muthulingam, 2019), and entity responsible for financing the inspection (Short et al., 2016).

According to three key articles identified and reviewed by Greenfield et al. (2007) and Klerks et al. (2013), previous studies focus exclusively on inspection outcome and lack consensus about the effects of announced and unannounced inspections. On the one hand, Klerks et al. (2013) and Ehlers et al. (2017) do not find any empirical evidence in the healthcare industry that inspection outcomes from announced and unannounced inspections at the same organizations are statistically significantly different. Short et al. (2016) also show that the control variable that indicates whether an inspection is announced was insignificant in the context of global supply chain. On the other hand, other studies provide evidence of the beneficial effect of announced over unannounced inspections. For example, Reske et al. (2007) demonstrate that the restaurants that received announced inspections in addition to routine unannounced inspections report better inspection outcomes (fewer violations) than the restaurants that received only routine unannounced inspections. Greenfield et al. (2012) similarly suggest that Australian hospitals are more likely to meet inspection outcome standards if subjected to 2-day notice announced inspections rather than unannounced inspections. Besides, based on the results yielded by analytical models, Kim (2015) demonstrates that announced inspections are superior to unannounced inspections in overall efficiency. Consequently, current studies do not fully comprehend the different aspects of announced and unannounced inspections. First, previous studies do not fully incorporate operational perspective. The authors do not go beyond articulating the relationship between inspection types, announced and unannounced inspections, and inspection outcome. Second, there is no empirical support for the beneficial effect of unannounced inspections, which contradicts the popular use of unannounced inspections in practice. Third, a coherent theory elucidating the underlying distinction between announced and unannounced inspections is presently lacking. Our study fills these important gaps in the current literature.

Table 3.2 Literature Review

Article	Industry sector	Research focus†	Independent variables	Dependent variables	(a)	(b)	(1)	(2)
Ball et al. (2017)	Medical device	Investigate the moderating role of inspector experience on the relationship between inspection outcome and future product recalls	Inspection outcome, inspector experience	Future recall (operational)	X			X
Scott (2018)	Road	Examine how work environment (bad weather) influences inspectors' productivity	Weather	Total inspections; inspection outcome	X			X
Bennett et al. (2013)	Vehicle test	Examine the relationship between competition and inspection stringency	Competition	Inspection outcome	X		NA	NA
Ibanez & Toffel (2019)	Restaurants & food	Study the relationship between inspection scheduling and inspection quality	Inspector's schedule, Inspector experience (e.g. prior outcome)	Inspection outcome	X		*	*
Reske et al. (2007)	Restaurants	Study how announced inspections affect food safety - announced inspections reduce critical violations	Announced and unannounced inspections	Inspection outcome	X		X	X
Short et al. (2016)	Various	Identify the factors that influence violation detection - inspection outcome is unaffected by inspection type (announced or unannounced)	Inspector experience, team composition	Inspection outcome	X		X	X
Ehlers et al. (2017)	Hospital	Study the effectiveness of announced and unannounced inspections in detecting violations - unannounced inspections are not more effective than announced inspections with respect to non-compliance detection	Inspection type (announced or not)	Inspection outcome	X		X	X
Greenfield et al. (2012)	Healthcare	Compare announced inspection and short-notice inspection (2 days or 1.5 hour before) - unannounced inspections detect more non-compliance than announced inspections	Inspection type (timing of announcement)	Inspection outcome	X		X	X

(a) Inspection body or Inspector's perspective (b) Inspected organization's perspective
(1) Announced inspection (2) Unannounced inspection
* Inspection announcement is not explicitly mentioned or controlled (routine and non-routine inspections are included in the data)
** Announced and unannounced inspections are included in the data, but interpretation is limited due to the empirical setting.
†This column includes research question from inspection studies, but the findings are further reported only if the study is relevant to announced and unannounced inspections.

Table 3.2 Literature Review (Continued)

Article	Industry sector	Research focus	Independent variables	Dependent variables	(a)	(b)	(1)	(2)
Dhanorkar et al. (2018)	Mfg.	Investigate whether punitive tactics (inspection) help firms in implementing supportive program	Inspection (type, outcome) Recommendation	Program implementation rate		X		X
Gray et al. (2011)	Pharma	Identify which factors influence quality risk (inspection outcome) in offshore and domestic plants	Employee skill, geographic and cultural distance	Inspection outcome		X		X
Levine et al. (2012	Various	Investigate how occupational safety inspection influences injury rate	Unannounced inspection	Injury rate		X		X
Lu & Wedig (2013)	Nursing homes	Investigate the relationship between geographic clustering and quality (inspection outcome)	Chain size	Inspection outcome		X		X
Gray et al. (2015)	Pharma	Study the effect of R&D colocation on conformance quality (inspection outcome) and identify moderators	Colocation, moderators (technology, knowledge, size)	Inspection outcome		X		X
Mani & Muthulingam (2018)	Oil	Examine relationship between organizational learning from inspection experience and environmental performance (inspection outcome)	Direct and indirect inspection experience	Inspection outcome		X	*	*
Anand et al. (2012)	Pharma	Study whether decay in operational routines is systematic and predictable	Time since last inspection, M&A	Inspection outcome		X	**	**
Klerks et al. (2013)	Nursing homes	Investigate the differing effects of announced and unannounced inspections on risk detection (inspection outcome) - unannounced inspections do not report more violations than announced inspections - while inspectors noted a major difference between two inspections, in interviews, managers stated that they do not perceive much difference	Inspection type (announced or not)	Inspection outcome	X	X	X	X
This paper	**Nursing homes**	**Study how announced and unannounced inspections differently affect the level and the variation in quality performance**	**Inspection type (announced or not)**	**Clinical quality**		**X**	**X**	**X**

(a) Inspection body or Inspector's perspective (b) Inspected organization's perspective
(1) Announced inspection (2) Unannounced inspection
* Inspection announcement is not explicitly mentioned or controlled (routine and non-routine inspections are included in the data)
** Announced and unannounced inspections are included in the data, but interpretation is limited due to the empirical setting.
†This column includes research question from inspection studies, but the findings are further reported only if the study is relevant to announced and unannounced inspections.

Attention-Based View

The attention-based view (ABV) of an organization offers a theoretical lens on how organizational attention influences performance. According to this theory organizational attention refers to "the socially structured pattern of attention by decision-makers within an organization" (Ocasio, 1997). Shepherd et al. (2017) define top managers' attention as "the focusing of time, energy, and effort on issues and answers" and classify it into transient and sustained attention. *Transient attention* refers to "a fleeting focus of time, energy, and effort on a particular task" and *sustained attention* refers to "a prolonged focus of time, energy, and effort on a particular task" (Shepherd et al., 2017).

Ocasio (1997) proposes in ABV that organizations (firms) are "systems of structurally distributed attention" and that organizational behavior results from allocating attention of decision-makers to the specific context and situation that organizations face. Specifically, Ocasio (1997) proposes three aspects of organization attention: focus of attention, situated attention, and structural distribution of attention. First, focus of attention indicates "what decision-makers do depend on what issues and answers they focus their attention on." Next, situated attention indicates "what issues and answers decision-makers focus on, and what they do, depends on the particular context or situation they find themselves in." Lastly, structural distribution of attention refers to "what particular context or situation decision makers find themselves in, and how they attend to it, depends on how the firm's rules, resources, and social relationships regulate and control the distribution and allocation of issues, answers, and decision-makers into specific activities, communications, and procedures." We delineate announced and unannounced inspections in the context of nursing homes using the ABV.

Attention to Announced Inspection

The top management team in a nursing home consists of administrators, directors of nursing, and medical directors. They coordinate staff, care, and services at the nursing home, and manage the external regulation and accreditation processes (Castle et al., 2009). They focus attention and resources on inspections because it leads to quality improvement and signals high quality to consumers, which improves organizational survival (JCAHO & JCR, 2007; Ruef &Scott, 1998; Su & Linderman, 2016; Wagner et al., 2012b; Westphal et al., 1997). In addition, they allocate their time and attention to inspections because failure to comply would damage the nursing home's reputation. However, top managers' attention is limited (Ocasio, 1997), and a number of organizational issues compete for their attention. Therefore, the level and the duration of their attention that they allocate to an inspection may depend on whether the inspection is announced or unannounced.

In an announced inspection, the organization receives notification of the inspection well in advance to the site visit. In addition, an inspection is a non-routine event and not part of the typical daily operating activities in a nursing home. Top managers have limited attentional resources for a non-routine event and other important issues compete for their attention (Hoffman & Ocasio, 2001; Ocasio, 1997). When an inspection is scheduled far in advance, managers may focus their attention on the immediate short-term daily operational needs such as managing staff problems. Thus, an announced inspection results in transient attention whereby the top managers will respond to an inspection if and only if the scheduled inspection is approaching and requires immediate preparations for the visit. According to some of our interviews of healthcare administrators as well as a case study conducted by TJC (JCAHO & JCR, 2006, 2007), the administrators indicated that preparing for an announced inspection is like "cramming for an exam in college," "everybody is scrambling around the facility in the months and weeks before the inspection," and "an intense period of time when we would do tons of work." During this period, managers allocate their attention and resources to the inspection in a temporary but intense manner.

When an inspection is announced, managers allocate transient attention through interactions and communication with the employees about the impending inspection. For instance, nursing homes need to post the upcoming inspection schedule at all entrances to alert staff members before the inspectors' visit, which creates an intended awareness and concern about the inspection (JCAHO, 2004; JCAHO & JCR, 2006). Such transient attention may last several weeks, because preparing for an inspection requires that all employees be familiar with the current standards and processes before the inspection visit.[10] After the inspection, attention may be reallocated to other emerging priorities since it is a scarce resource. Administrators also noted that after the inspection everyone breathed a sigh of relief and went back to their normal way of doing things, indicating that attention to the accreditation standards might not be sustained over time (JCAHO & JCR, 2006, 2007). Specifically, one administrator said "We monitored for a little bit of time, but we'd stop monitoring (after the inspection)." In sum, under the announced inspection strategy, top managers allocate high levels of attention to the accreditation criteria for a limited time.

Attention to Unannounced Inspection

In the unannounced inspection strategy, the nursing home receives an unscheduled inspection that is not known until the morning of the site visit. This inspection is also a non-routine event, but top managers cannot allocate their attention away from the accreditation criteria since

---

[10] Retrieved from https://www.centrahealthcare.com/calling-all-rehab-professionals-how-to-prepare-for-a-joint-commission-survey/ and https://www.jointcommission.org/assets/1/18/LTC_10_steps_to_Accred_08.pdf

they cannot plan in advance for the site visit. Managers need to always be diligent to the accreditation criteria since they cannot plan for the inspection in advance and could potentially receive a negative inspection outcome. To avoid a negative inspection outcome, managers are likely to allocate a sustained level of attention to the accreditation criteria. Interestingly, our conversation with healthcare administrators that experienced unannounced inspections substantiate this thinking: they indicated that unannounced inspections are less disruptive than announced inspections because staff can concentrate on their job without looking at the inspection schedule. One administrator stated that his staff retained the same focused intensity on the criteria after the inspection as they did before the inspection (JCAHO & JCR, 2006, 2007).

Ocasio and Joseph (2017) suggest that sustained attention engenders both formal and informal interaction and communication on the selected issue. Sustained attention on an issue results in getting incorporate into routine meetings, reports, and administrative protocols (Ocasio, 1997). In practice, top managers at nursing homes prepare a checklist, distribute leadership notebook, held periodic quality resource councils, and educated staff to increase employee engagement in continuous readiness.[11] Since unannounced inspections are not viewed as scheduled events, but they motivate the organization to continuously meeting their quality and service standards. Thus, top managers will maintain sustained levels of attention on the accreditation criteria.

Attention and Quality

While the previous literature argues that inspections influence quality performance (Anand et al., 2012; Ball et al., 2017; Gray & Shimshack, 2011; Mani & Muthulingam, 2019), we know little about the relationship between inspection mode (announced/unannounced) and quality performance. However, the ABV implies that unannounced inspections will likely cause managers to maintain a sustained level of attention on accreditation. But, an announced inspection has a greater immediate effect than an unannounced inspection, because a prior notice leads to intense transient attention of top managers and provides time and opportunity to enhance the immediate quality performance. Instead, unannounced inspections lead to a sustained attention that may eliminate the ramp-up approach. These arguments suggest the following hypotheses:

> *H1. a) Announced and b) unannounced inspections increase the immediate quality performance.*

[11] Transcript of unannounced survey telephone conference call on March 9, 2006. Retrieved from http://www.jointcommission.org:80/NR/rdonlyres/8220A9E6-5278-4FAE-A7BB-FB14DAB631AE/0/unann_surv_transcript.pdf via wayback machine (https://archive.org/web).

H1c. Announced inspections have a <u>greater effect on immediate quality performance</u> than unannounced inspections.

However, unannounced inspection visits are more likely to exert a persistent effect on quality performance relative to the announced inspection visits. For instance, announced inspections that call for immediate attention can prompt the facility managers to exhibit an opportunistic behavior, while reverting to the normal practices after the inspection without making any systemic or permanent changes. In contrast, an unannounced inspection draws sustained attention and mandates conformance to standards on an ongoing basis, making it more likely to result in continuous improvements in quality performance. This suggests the following hypotheses:

H2. a) Announced and b) unannounced inspections increase sustained quality performance.

H2c. Announced inspections have a smaller effect on sustained quality performance than unannounced inspections.

3.4 Data and Variables

Sample and Data Collection

The data comes from the CMS, Long-Term Care Focus, and TJC. We first obtained a full list of Medicare or Medicaid certified nursing homes from the CMS Nursing Home Compare. We then retrieved the nursing home name, address, and characteristics, including ownership, hospital affiliated status, and quality measure items that are reported and updated on a quarterly basis. Next, we obtained other relevant nursing home information from the CMS Provider of Services (POS), CMS Cost Report, and Long-Term Care Focus datasets. The POS dataset had information on the number of beds, nursing home chain membership (network affiliation), and staffing information. The Cost Report had financial information such as operating revenue and expenses. The Long-Term Care Focus contains important resident variables that CMS Online Survey, Certification and Reporting (OSCAR) and Certification and Survey Provider Enhanced Reports (CAPSER) have offered. We thus obtained facility-level acuity index and payer mix information from Long-Term Care Focus. Then we merged the data using Medicare provider number, which is a unique identifier across data sources. Finally, we obtained nursing home name, address, accreditation program, inspection date, and accreditation outcome by scraping from the TJC Quality Check website (http://www.qualitycheck.org/). We combined the TJC data with the previous dataset by matching the nursing home name and address because TJC data does not contain the Medicare provider number. This process resulted in 834 TJC accredited nursing homes with a Medicare provider number after excluding prisons, VA facilities and non-Medicare or non-Medicaid facilities.

We use nursing home quarterly data in the analysis. Quarterly data is the shortest time interval in our dataset, and using yearly data may not capture fluctuations in quality performance caused by inspections. Quarterly data also enables us to examine both immediate and sustained effects of inspections. However, we use data range from 2013 to 2016 for several reasons. First, we do not have data from 2006, when the TJC regime changed from announced to unannounced inspections. We note that more reliable and accurate quality measures in Nursing Home Compare were released after implementing MDS 3.0 in late 2010.[12] Next, the current data range allows us to provide a reliable estimate of announced and unannounced inspections conducted since 2013. It is critical to know the current and previous inspection dates and types (e.g., initial or recurring inspection) to identify whether the current inspection is announced or unannounced. Since TJC provides a comprehensive accreditation history from 2009 and the inspection interval can be up to 39 months, 2013 is the appropriate starting point of identification.[13]

Dependent Variables

The dependent variable of nursing home quality comes from Nursing Home Compare and consists of multiple items that measure quality. These quality items represent overall quality rather than inspection outcome (e.g., non-compliance during an inspection), and they do not come from the inspectors or the accreditation decision.[14] Therefore, the items do not directly influence the accreditation outcome. However, these items are important because the set of quality measures are converted to CMS nursing home star ratings, where potential consumers can access this quality performance rating of a nursing home. We selected these eleven items (see Table A1) that have been included in determining the nursing home star rating between 2013 and 2016. These items assess how each nursing home addresses its residents' physical and clinical needs. The items have good accountability, as the Research Triangle Institute and the National Quality Forum demonstrated the validity, reliability, and reportability for each one.[15] We first convert a value of each item to an 100-point scale following the scoring method in the nursing home star rating (CMS, 2012). For each item, we assign 100 points for top performers. Then, we assign 1 to 99 points for

[12] https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/NHQIMDS30.html
[13] We note that TJC updated the interval of unannounced inspections to 18–36 months in 2011, while previously it was 18–39 months (GAO, 2011).
[14] Confirmed through interviews with an associate director at TJC and a previous healthcare administrator.
[15] https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/index.html

remained nursing homes based on the updated national percentile after eliminating top performers. The only exception is the activities of daily living (ADL) item in Table A1. We use the state-specific quantile for the ADL measure, because ADL can be influenced by state Medicaid policies. Following the CMS manual, we assign 100 points for the best performers, 20 points for the worst performing quantile, and 40, 60, and 80 points for the second, third, and fourth quantile, respectively (Centers for Medicare and Medicaid Services 2012). For the nursing homes in the states with fewer than five facilities, we use the quantile from national distribution, as the manual suggests.

We categorize the eleven items into three groups, all residents, long-stay residents, and short-stay residents, instead of considering individual item respectively. The residents are denoted as long-stay residents if they stayed in a nursing home for a period longer than 100 days and are short-stay residents otherwise. Long-stay quality items are collected only from long-stay residents and short-stay quality exclusively from short-stay residents by CMS. There are eight items for long-stay residents and three items for short-stay residents in Table A1. To create measures for the dependent variables, we first construct the most comprehensive measure, all resident quality, by adding up all 11 items in nursing home star rating in Table A1. Next, we compute a measure of long-stay (short-stay) quality by aggregating long-stay (short-stay) resident items in the nursing home star rating.

It is also important to note that the computed quality measures represent relative quality not absolute quality, similar to Su and Linderman (2016) who used this notion to highlight that high-quality stands for an organization's overall quality relative to its competitors. This concept of relative quality corresponds to the achievement score in nursing home Value Based Purchasing program, which forms the basis of CMS' pay for performance program starting in 2019.[16]

Independent Variables

To examine the immediate effects of inspection in H1, we construct two dummy variables that indicate whether a nursing home received an announced or unannounced inspection visit in a given corresponding quarter. In other words, the announced visit variable takes the value of 1 if a nursing home receives an announced inspection visit at quarter *t*. For non-TJC-accredited nursing

[16] https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Other-VBPs/SNF-VBP.html

homes, the value of this variable is always 0. The coding of these variables including exceptional conditions was verified extensively.[17]

To investigate the sustained effects in H2, we create dummies for relative timing of inspection entry, up to two-quarter leading and lagged variables for both inspection visit indicators. The coding is similar to and consistent with other empirical studies (see for instance, Autor (2003) and Hydari et al. (2019)). We do not consider periods more than two quarters because the post-effect of the current inspection might be contaminated by the pre-effect of the next inspection if a nursing home has a minimum six-quarter interval between inspections.

Matching Variables

Ownership. Ownership may predict quality level and thus the likelihood of applying for the TJC accreditation program (Chesteen et al., 2005; Wagner et al., 2012a). For-profit nursing homes would likely apply for the accreditation program because the TJC accreditation signals superior quality to potential customers who seek better care.[18] However, for-profit nursing home management may be also interested in cutting costs instead of investing resources into the TJC accreditation program, which requires higher standards. Conversely, governmental nursing homes are more likely to seek to deliver higher level of quality and safety through TJC accreditation. To control for nursing home ownership, we use two ownership dummies variables to describe three ownership groups: for-profit, not-for-profit, and governmental nursing homes as a baseline group. This classification has been used in several other studies (Afendulis et al., 2016; Grabowski et al., 2016, 2004).

Chain membership. To control for the effect of chain membership (network affiliation), we use a dichotomous variable indicating whether a nursing home is owned by a multi-facility organization, a chain. Members of nursing home chains can achieve greater economies of scale and higher quality (Lu & Wedig, 2012; Wagner et al., 2012a). We note that some large chains, such as Lexington Health Network, participated in TJC accreditation as a set.

Location. Nursing homes in urban areas are more likely to be found accredited than those located in rural areas (Wagner et al., 2012a). Thus, to control for location, we create a dummy variable and assign it the value of 1 if a nursing home is located in an urban area.

---

[17] We verified them by reviewing hospital affiliation status, accreditation program type, and accreditation history from 2010 to 2016. Further discussion of exceptions is provided in the Robustness Check section.

[18] TJC has touted the benefits of its accreditation, including gaining competitive advantage, signaling high quality to consumers, and ensuring greater quality and safety. (https://www.jointcommission.org/assets/1/6/2009_LTC_Overview_Combo_10_30_09.pdf)

TJC state. Thirteen states recognize TJC's nursing home accreditation program while the remaining states do not.[19] For instance, the Department of Labor and Employment in Colorado recommends TJC accreditation for long-term residential services at nursing homes, and the Department of Community Health in Georgia considers TJC accreditation as evidence of established compliance with the departmental requirements. To control for these effects, we create a dichotomous variable indicating whether a nursing home is located in the state that recognizes TJC accreditation.

Size. Large nursing homes may have higher level of quality and are more likely to apply for accreditation (Lu & Wedig, 2012; Wagner et al., 2012a). Extant literature suggests that nursing homes can be classified into three groups depending on the size: small (with 75 beds or fewer), medium (76–125 beds), and large (126 beds or more) (Afendulis et al., 2016; Grabowski et al., 2016). To control for nursing home size, we use two dummy variables representing small, medium, and large size, treating small as the baseline level.

Payer mix. Payer mix is a factor affecting financial resources because nursing homes generate greater margins from Medicare and private-pay residents than from Medicaid residents (Lu et al., 2018; Nyman, 1993). To control for the payer mix, we created two continuous variables denoting the percentage of residents in a nursing home whose payer is Medicare or Medicaid, respectively (Afendulis et al., 2016; Grabowski et al., 2016).

Acuity index. To control for an average intensity of nursing care required by residents in a nursing home, we use an acuity index. We obtain the acuity index from Long-Term Care Focus, where this measure is constructed by considering the ratio of residents with respect to various levels of activities of daily living (ADL) assistance and special treatment (Cowles, 2007).

Operating margin. Operating margin is one of the profitability indicators that represent the nursing homes' ability to generate financial return (Bowblis, 2011; Pradhan et al., 2013). Operating margin is possibly associated with TJC accreditation because only nursing homes that are financially viable can apply for the accreditation program.[20] To control for the operating margin, we create a continuous variable by computing the formula, whereby each item is extracted from the Cost Report.

$$\textit{operating margin} = (\textit{operating revenue} - \textit{operating expense}) / \textit{operating revenue}$$

[19] https://www.jointcommission.org/state_recognition/state_recognition.aspx

[20] The typical three-year accreditation fee was $9,700 in 2009 that can vary depending on the nursing home size (https://www.jointcommission.org/assets/1/6/2009_LTC_Overview_Combo_10_30_09.pdf).

Missing Variables and Final Sample

Figure 3.1 summarizes the steps taken to prepare the dataset for analyses. We first remove the nursing homes with missing variables of payer mix, acuity index, and operating margin. This resulted in 731 TJC accredited nursing homes from the 13,326 Medicare (or Medicare) certified nursing homes between 2013 and 2016. Next, our sample size was reduced further to 452 TJC accredited nursing homes from the sample of 4,927 nursing homes for which all dependent variables were available. Specifically, CMS publicly reports each quality measure item only if a nursing home has a sufficient number of residents. For instance, long-stay resident measures should mandate at least 30 long-stay residents and short-stay resident measures require at least 20 short-stay residents in a nursing home (Smith et al., 2012). Based on the data availability of these items, CMS labels nursing homes as long-stay facility, short-stay facility, or both. Long-stay facilities are the nursing homes that successfully report all long-stay quality items and short-stay facilities are the nursing homes with all reported short-stay quality items (CMS, 2012). Given 452 TJC accredited nursing homes that belong to both long-stay and short-stay facility, we could compare quality from long-stay residents and short-stay residents. Third, we consider 337 TJC-accredited nursing homes that received only one inspection, either announced or unannounced, from 2013 to 2016: We have 225 nursing homes that received only one announced inspection and 112 nursing homes that received only one unannounced inspection. In the robustness section, we also discuss nursing homes that received two inspections (e.g., two announced inspections, two unannounced inspections, or one announced and one unannounced inspection) during the 2013−2016 study period.[21]

Table 3.3 reports descriptive statistics for nursing homes whose quarterly quality measures items are available. Table A2 shows that these nursing homes are more likely to be large and located in an urban area, with relatively lower operating margin, and have residents with higher hospitalization risk (acuity index) than an average nursing home.[22] This is consistent with our sample reduction approach because we have removed the nursing homes that serve fewer than 30 long-stay residents or 20 short-stay residents.

[21] It should be noted that 115 nursing homes received two inspections by TJC between 2013 and 2016. Specifically, 78 nursing homes received two announced inspections, 21 received two unannounced inspections, and 16 received one announced and one unannounced inspection.

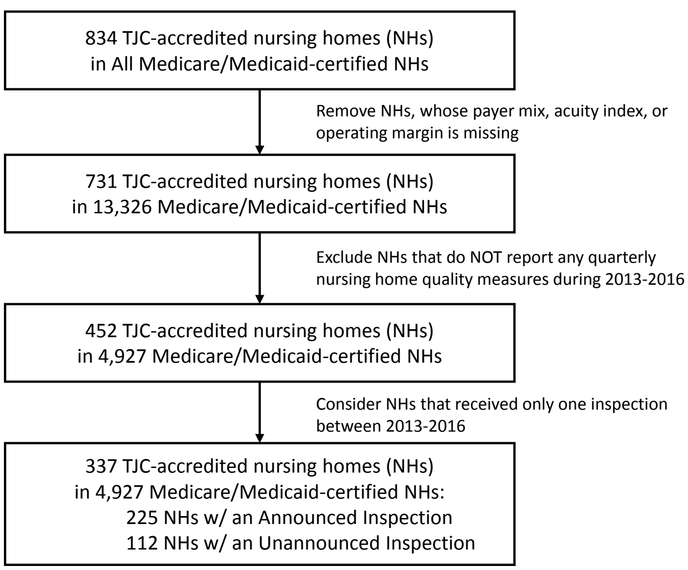
[22] We did not use imputation strategy for missing dependent variables because many values were missing from our quarterly-based dataset. If we implemented the state-average imputation for the nursing homes, whose number of residents is close to the reporting margin, the imputation would create abrupt changes and would result in a quality level overestimate for those nursing homes that actually have lower quality.

Table 3.3 Summary Statistics and Data Sources

Variable	Description (*n* = 4,927; TJC-accredited = 452; non-TJC-accredited = 4,475)	Mean	SD	Source
Quality Measure (QM) at time = 1				
All Quality	Sum of 11 QMs in nursing home star-rating	660.2	119.9	NH Compare
Long-stay Quality	Sum of 8 QMs of long-stay residents in nursing home star-rating	472.6	93.6	NH Compare
Short-stay Quality	Sum of 3 QMs of short-stay residents in nursing home star-rating	187.7	51.5	NH Compare
Independent Variables				
Announced Visit†	1 if a NH received announced visit by TJC, otherwise 0 (4.5%)			TJC
Unannounced Visit†	1 if a NH received unannounced visit by TJC, otherwise 0 (2.3%)			TJC
Matching Variables				
Ownership†	Profit (74.8%), non-profit (22.1%), government (3.1%)			Provider of Services
Chain†	Whether a NH is owned by a multi-facility organization (60.2%)			Provider of Services
Location†	Whether a NH is located in urban area (83.0%)			Provider of Services
TJC State	Whether a NH is in the state that recognizes TJC accreditation (33.3%)			TJC
Size	number of beds	147.3	63.7	NH Compare
Medicare (%)	% of Medicare resident in a NH	16.1	9.0	Long Term Care Focus
Medicaid (%)	% of Medicaid resident in a NH	61.3	15.1	Long Term Care Focus
Acuity Index	Average intensity of nursing care required by residents in a NH	12.4	1.0	Long Term Care Focus
Operating Margin (%)	[(operating revenue + operating expense)/operating revenue]*100	-2.3	27.4	Cost Report

† Only percentage values are reported for categorical variables

Figure 3.1 Steps of Finalizing the Sample



3.5 Analysis and Results

Empirical Strategy

We adopt a matching method for TJC-accredited nursing homes, which serves as the treatment group, and non-TJC-accredited nursing homes comprising the control group. This method can address selection bias by balancing the distribution of characteristics that are potentially associated with treatment and outcome between treated and controlled nursing homes (Blackwell et al., 2009). It is important to note that a nursing home's application for the TJC accreditation is always voluntary and thus non-random. Among various matching options, we opt for coarsened exact matching (CEM) over the more popular propensity score matching (PSM) method for several reasons. First, PSM can cause a random matching problem whereby the paired treatment and control have imbalanced covariates even with similar propensity score. Next, CEM is more appropriate when both continuous and discrete matching variables are employed (King and Nielsen 2019). Since CEM considers a coarsened range of variables rather than exact covariate values, CEM approximately provides a fully blocked experimental design (Blackwell et al., 2009). We use the following variables from previous nursing home studies to find a match: ownership, chain, membership, location, size, payer mix, and acuity index (Afendulis et al., 2016; Grabowski et al., 2016; Wagner et al., 2012a). We further include two matching variables that related to TJC accreditation - TJC state and operating margin.

We follow existing literature to deal with each continuous variable. For payer mix, we replace continuous variable with dummy variables following the previous literature to minimize the curse of dimensionality (Afendulis et al., 2016; Grabowski et al., 2016). In contrast, we retain the acuity index and operating margin as continuous variables. For operating margin, we specify a set of pre-defined cut-points (e.g., 0%, 5%) in CEM where lower than 0% is considered as low operating margin, 0–5% as moderate operating margin, over 5% as high profitability (Cadigan et al., 2015).

For 337 TJC-accredited nursing homes that are eligible for inclusion into both long-stay and short-stay facility groups, the implementation of one-to-one without replacement CEM leads to 331 TJC-accredited nursing homes and 331 non-TJC-accredited nursing homes. Table A2 shows that TJC-accredited nursing homes are more likely to be for-profit nursing homes located in urban area, with a chain membership and relatively higher percentage of Medicare patients. However, we confirm that TJC-accredited and non-TJC-accredited nursing homes are well balanced with respect to these characteristics after matching.

To examine H1 we conduct panel regression analyses with nursing homes and unit time fixed effects. We include a lagged dependent variable term because past quality level may persist in the future; that is, the current care and service is the function of past care and service, which may be modified by an inspection. We develop announced visit and unannounced visit models separately by implementing two distinct equations. Specifically, we first consider TJC-accredited nursing homes that received one announced inspection and non-TJC-accredited nursing homes. Next, we analyze differences between TJC-accredited nursing homes that received one unannounced inspection and non-TJC-accredited nursing homes.

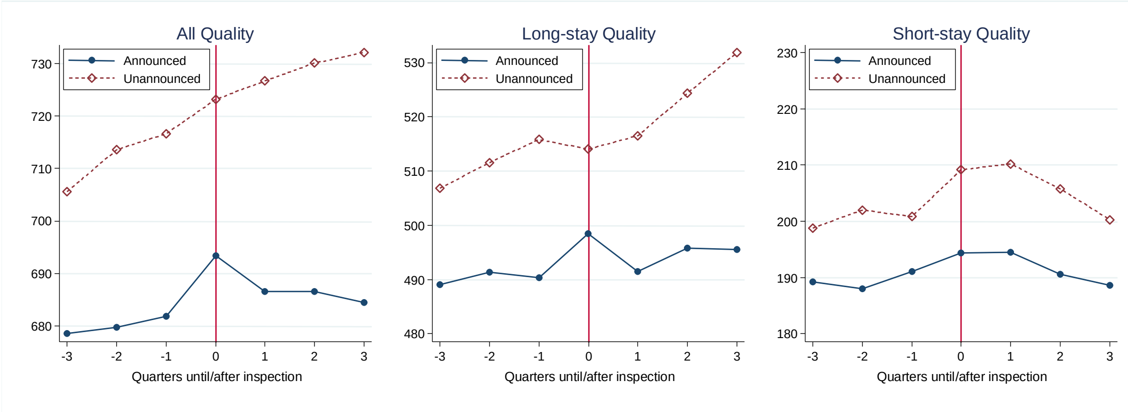
Next, to investigate sustained effects of inspection visits postulated in H2, we add up to two-quarter lags and leads of the announced inspection visit indicator () and the unannounced inspection visit indicator () in the above equations.[23] This relative timing model has been used in many studies to investigate the causal and sustained effects (Autor, 2003; Dhanorkar, 2019; Hydari et al., 2019). This model has several advantages. First, it allows us to examine the changes to inspection visit effect over time, before and after the inspection visit. Specifically, lead variables detect pre-treatment effects and lagged variables capture post-treatment effects. Second, we can test parallel trends assumption in this model, while a traditional difference-in-difference model fails to address the issue of different pre-trends between treatment and control groups (Autor, 2003; Dhanorkar, 2019; Hydari et al., 2019). For instance, unlike to our approach, a traditional difference-in-difference model should consider an entire period before the event as a baseline. Lastly, we can evaluate the causality concern by investigating coefficients of lead variables. The causality is supported if results suggest no pre-treatment effects.

Results

Figure 3.2 shows the changes in the quality levels near the inspection visit. The y axis indicates the quality score from all residents, long-stay residents, and short-stay residents, respectively, and the x axis indicates the time relative to the quarter of the inspection visit. Specifically, the reference time (t = 0) is the quarter when a nursing home received either announced or unannounced inspection visit. Overall, we observe notable patterns indicating that the quality level increases as the inspection visit approaches. However, the quality level peaks at the time of the announced inspection visit and then takes a slight dip. In contrast, the level persists or rises slightly after the unannounced inspection visit.

[23] Our notation follows Hydari et al. (2019). For instance, announced visit follows: $Quality_{it} = \alpha_i + \beta_{-2}A_{it+2} + \beta_{-1}A_{it+1} + \beta_0 A_{it} + \beta_1 A_{it-1} + \beta_2 A_{it-2} + \gamma Quality_{it-1} + \sum_3^{16} \delta_t Time_t + \mu_t + \varepsilon_{it}$. where A_{it} is $Announced\ Visit_{it}$. Lead variables detect pre-treatment effects and lagged variables capture post-treatment effects.

Figure 3.2 Quality Level Changes Near the Time of Inspection Visit



Notes. Figure 3.2 displays the changes in the quality levels near inspection visits. The *y* axis indicates the quality score from all residents, long-stay residents, and short-stay residents, respectively. The *x* axis represents the time passage relative to the quarter of the inspection visit. Specifically, the reference time () is the quarter when a nursing home received either announced or unannounced inspection visit.

We first examine H1 for the presence of immediate effects of announced and unannounced inspections on quality. Table 3.4 shows the effects of both inspections on three dependent variables: quality performance of all residents (column 1), long-stay residents (column 2), and short-stay residents (column 3). The results in Table 3.4 provide partial support for H1a. We find that an announced inspection visit increases immediate quality performance for all resident (b = 11.206, $p < 0.05$) and long-stay resident quality (b = 8.094, $p < 0.05$), but the results are not significant for short-stay residents (b = 2.903, $p > 0.10$). Next, the analysis indicates that an unannounced inspection visit has a positive and significant impact on short-stay resident quality (b = 8.516, $p < 0.01$) but the results are not significant for all-stay and long-stay resident quality. These results partially support H1b. To examine H1c, we compare regression coefficients across the two models related to announced and unannounced inspection visits, following the procedure introduced by Clogg et al. (1995). We find a marginally significant difference in long-stay resident quality (Δb = 8.527, $p < 0.10$), indicating that the immediate quality increase due to an announced inspection visit is greater than an unannounced inspection visit, which supports H1c.

Table 3.4 Immediate Effects of Inspection Visits

	(1) Announced Inspection			(2) Unannounced Inspection		
	All residents Quality	Long-stay Quality	Short-stay Quality	All residents Quality	Long-stay Quality	Short-stay Quality
	1a	2a	3a	1b	2b	3b
Inspection Visit	11.206*	8.094*	2.903	7.671	-0.433	8.516**
	(4.677)	(3.732)	(2.288)	(5.905)	(4.963)	(2.817)
DV (t-1)	0.516**	0.470**	0.489**	0.526**	0.473**	0.497**
	(0.012)	(0.012)	(0.010)	(0.013)	(0.014)	(0.012)
Constant	349.940**	275.167**	103.434**	344.446**	274.804**	102.789**
	(8.693)	(6.607)	(2.527)	(9.640)	(7.522)	(2.871)
Nursing FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Nursing Homes	552	552	552	441	441	441
Observations	8,280	8,280	8,280	6,615	6,615	6,615
R^2	0.282	0.241	0.241	0.299	0.249	0.250

Clustered Standard Errors in parentheses; ** p<0.01, * p<0.05, †p<0.1.

Table 3.5 Sustained Effects of Inspection Visits

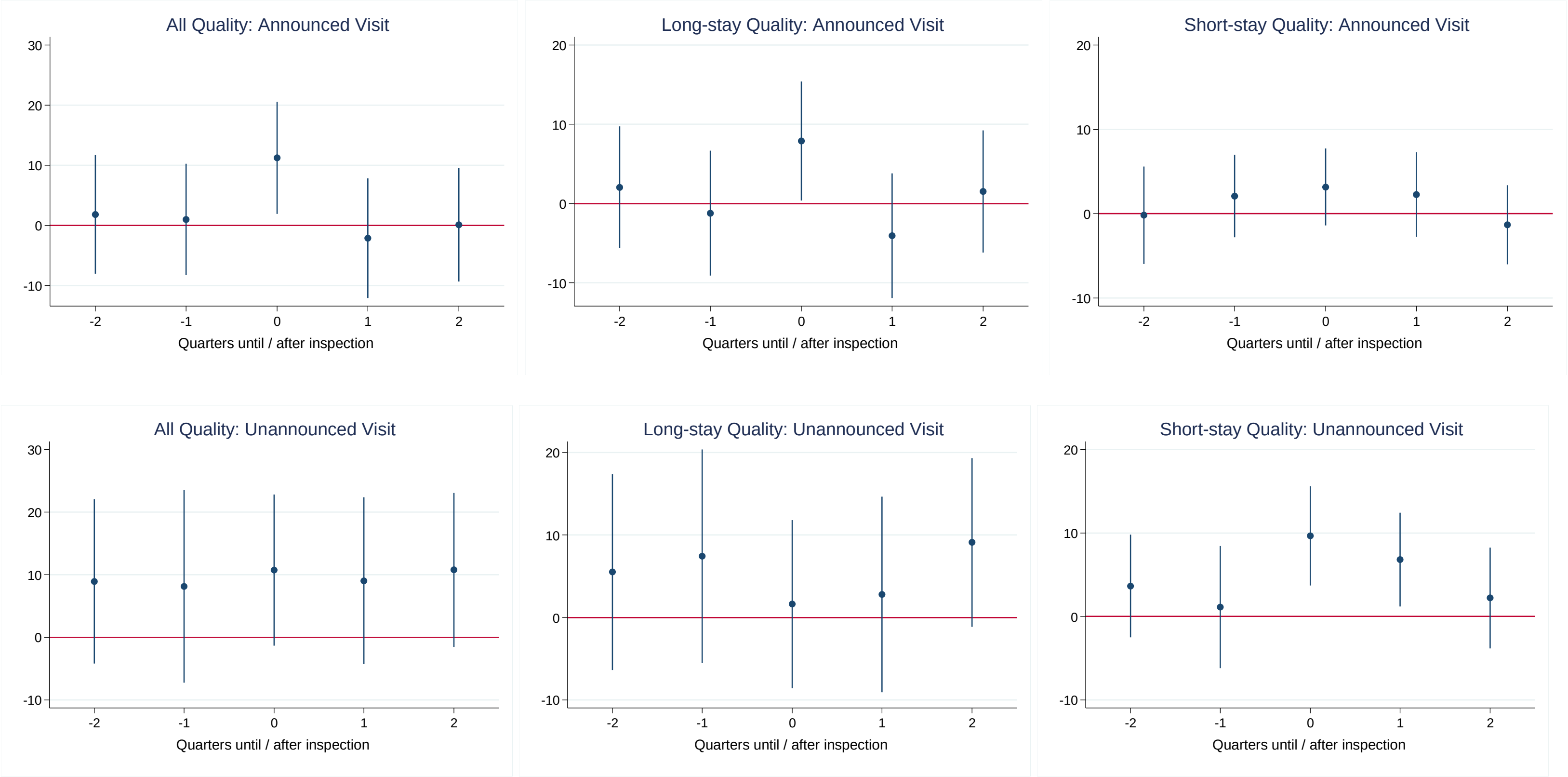
	(1) Announced Inspection			(2) Unannounced Inspection		
	All residents Quality	Long-stay Quality	Short-stay Quality	All residents Quality	Long-stay Quality	Short-stay Quality
	1a	2a	3a	1b	2b	3b
2 Quarters Prior	1.811	2.050	-0.183	8.940	5.505	3.646
	(5.027)	(3.915)	(2.942)	(6.677)	(6.046)	(3.135)
1 Quarter Prior	0.988	-1.203	2.095	8.118	7.417	1.112
	(4.712)	(4.014)	(2.497)	(7.816)	(6.588)	(3.722)
Inspection Visit	11.248*	7.901*	3.172	10.732†	1.619	9.664**
	(4.750)	(3.815)	(2.318)	(6.137)	(5.183)	(3.022)
1 Quarter After	-2.135	-4.056	2.262	9.040	2.791	6.800*
	(5.064)	(3.997)	(2.561)	(6.778)	(6.035)	(2.857)
2 Quarters After	0.100	1.531	-1.319	10.777†	9.096†	2.213
	(4.811)	(3.924)	(2.394)	(6.253)	(5.199)	(3.070)
DV (t-1)	0.516**	0.470**	0.489**	0.525**	0.472**	0.496**
	(0.012)	(0.012)	(0.010)	(0.013)	(0.014)	(0.012)
Constant	349.904**	275.096**	103.405**	344.999**	274.991**	102.949**
	(8.693)	(6.606)	(2.529)	(9.648)	(7.513)	(2.878)
Nursing FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Nursing Homes	552	552	552	441	441	441
Observations	8,280	8,280	8,280	6,615	6,615	6,615
R^2	0.282	0.242	0.242	0.299	0.249	0.250

Clustered Standard Errors in parentheses; ** p<0.01, * p<0.05, †p<0.1.

Next, Table 3.5 shows the sustained effects of announced and unannounced inspections for the three dependent variables—quality variables based on all resident (column 1), long-stay resident (column 2), and short-stay resident (column 3). We also include leading and lagging variables for announced and unannounced inspections in Table 3.5. First, we examine H2a. The results show significant increases in all resident quality ($b = 11.248$, $p < 0.05$) and long-stay resident quality ($b = 7.901$, $p < 0.05$) at the time of the inspection visit. However, no post-inspection effects are revealed. We note that the previous result related to H1a is still supported even after we incorporate leads and lags of the inspection visit indicators. Next, we investigate H2b which argues for a sustained effect of an unannounced inspection visit. In contrast to the announced inspection visit, the results show sustained effects of unannounced inspection visits. First, unannounced visits increase all resident quality measures. The positive effects are marginally significant at the time of the inspection visit ($b = 10.732$, $p < 0.10$) and two quarters after the inspection visit ($b = 10.777$, $p < 0.10$). Next, for long-stay resident quality, we observe a marginally significant effect two quarters after the inspection ($b = 9.096$, $p < 0.10$). Finally, the effect of an unannounced inspection visit on short-stay resident quality is significant at the time of the inspection visit ($b = 9.664$, $p < 0.01$) and one quarter after the visit ($b = 6.800$, $p < 0.05$). Overall, these results support H2b, suggesting that unannounced inspection visits increase sustained quality performance.

To examine H2c, we compare coefficients of announced and unannounced inspection visits. Figure 3.3 gives the plot of the coefficient estimates of leads and lags of inspection visit indicators and their 95% confidence intervals. This figure suggests that the announced inspection effect exhibits a greater fluctuation over time than the unannounced inspection effect. We use the approach previously proposed by Clogg et al. (1995) to compare these coefficients between announced and unannounced inspections. The results support H2c, indicating that the sustained effects of announced inspections on all resident quality are significantly smaller than those of unannounced inspections both one quarter after the visit ($\Delta b = -11.175$, $p < 0.10$) and two quarters after the visit ($\Delta b = -10.677$, $p < 0.10$).

Figure 3.3 Effects of Leads and Lags of Inspection Visits on Quality



Notes. The *y* axis indicates coefficient estimates on quality. Dots represent estimates of the difference between treatment and control groups with respect to relative time of inspection and vertical lines display 95% confidence intervals of these estimates. The *x* axis represents the time passage relative to the quarter of the inspection visit. Specifically, the reference time (t = 0) is the quarter when a nursing home received either announced or unannounced inspection visit.

Robustness Checks

We conduct several robustness checks for bias correction in the dynamic panel model, unobserved effects, sensitivity, re-matching, and placebo tests. First, we perform a bootstrap-based bias correction procedure for the dynamic panel model (Everaert & Pozzi, 2007). Previous studies indicate that fixed effect estimators may be inconsistent when the number of cross-sectional observations is large but the number of time periods is small or fixed (Nickell, 1981). Some notable solutions are based on generalized method of moments (GMM) approaches, including difference GMM (Arellano & Bond, 1991) and system GMM (Blundell & Bond, 1998). However, these GMM estimators raise the issues of weak instruments and poor small sample property (Bun & Windmeijer, 2010). The bootstrap-based bias correction can mitigate these concerns through easier implementation that does not require any decisions regarding instrument selection (Everaert & Pozzi, 2007). Our bootstrap procedure employs 250 (wild) bootstrap iterations with burn-in initialization, which allows general heteroscedasticity and does not require an assumption regarding the initial condition of the fixed or normal distribution. The results confirm that the outcomes remain consistent after correcting for possible biases in the dynamic panel, while the p-values decrease slightly (see Appendix B, Table 3).

Next, we address the concern regarding unobserved effects. We first focus on the presence of potential seasonal patterns in nursing homes. The number of residents in a nursing home peaks in winter and reaches a low point in summer (according to the description by the Long-term Care Focus). To control for this pattern, we replace the fourteen-quarter dummies with three-year dummies and three-quarter dummies. We verify that our results hold after replacing the time fixed effects.

Another concern related to unobserved effects is that we did not consider regular unannounced inspections by state survey agencies (SA), which are monitored by CMS.[24] We note that these inspections are slightly different from the TJC inspections in several respects. First, while only nursing homes that voluntarily apply for TJC accreditation would receive a TJC inspection, state-survey agencies' inspections are mandated for all Medicare or Medicaid certified nursing homes. Next, TJC accreditation is subject to annual accreditation and on-site inspection fees, but state-survey agencies' inspections are free of charge. Lastly, administrators in healthcare organizations perceive TJC inspections as accommodating, collaborative, collegial, and educational. In contrast, state-survey agencies' inspections are inspection-oriented and more

[24] State survey agencies have used unannounced inspections following the CMS State Operations Manual 2700A. Staffs of the Survey and Certification (S&C) office at CMS confirmed that the unannounced inspection process started around 1980.

stressful to staff members in organizations.[25,26] Apparently, state survey agencies' inspections do not affect TJC accreditation and inspection, because they are mandated by state governments for institutions that participate the Medicare or Medicaid program. To address any concerns that may arise from state survey agencies' inspection effects, we incorporate a dichotomous variable in our model indicating whether a nursing home received an unannounced inspection from the state government agency in the quarter. The results are robust after incorporating unannounced inspections by state government agencies (see Appendix B, Table 4). We observe non-significant effects of governmental unannounced inspections on quality performance, while the effects of TJC's inspections are statistically significant.

We also check sensitivity of the results by employing other measures for dependent variables, decomposing inspections into more granular categories, and investigating heterogeneity in the inspection effects. We first investigate whether the results are robust under alternative measures of dependent variables. We previously constructed quality performance measures for all residents, long-stay residents, and short-stay residents from the items listed in CMS Five-Star Quality Rating System. Now, we reconstruct quality measures by additionally including the items that are not listed in the star rating but are included in Nursing Home Compare, such as vaccination and mental care items. The results show stronger statistical significance, supporting the previous results.

Next, we classify inspections into more granular categories to check for sensitivity. Although we previously distinguished between announced and unannounced inspections, announced inspections can be further disentangled based on the inspection history and interval between the day of notice and the actual visit. TJC conducts 30-day announced inspections (n = 76) when a nursing home apply the first TJC accreditation and recurring 7-day announced inspections (n = 145) under exceptional conditions such as when the accreditation program was retired (see Appendix B, Table 5). We note that both the first 30-day announced inspection and the recurring 7-day announced inspection only lead to an immediate increase in quality performance (see Appendix B, Table 6), while there are no sustained effects. However, we cannot fully separate the confounding effects of first announced versus recurring announced inspections and 30-day versus 7-day announced inspections independently. The results indicate that a negative post-treatment effect of the (first) 30-day announced inspection and a positive immediate effect of the (non-first) 7-day announced inspection. This disparity suggests that 7-day announced inspections attract more intense level of transient attention than do 30-day announced inspections, with a less pronounced

[25] TJC's teleconference about unannounced survey on March 9, 2006.
[26] Interview with a previous administrator in healthcare organizations held on May 25, 2017.

decaying effect. However, since 2015, TJC no longer conduct 7-day announced inspections as a part of the TJC nursing home accreditation program.

Next, we investigate heterogeneity in the inspection effects. Although inspections may lead to an increase in quality performance, we do not know whether inspection effects are much greater in the nursing homes with lower baseline quality. We suppose that the relative quality of top performers may not fluctuate, because their performance is always robust irrespective of any external shocks, including announced and unannounced inspections. In contrast, bottom performers may have more room for improvement. To examine this heterogeneity effect, we conduct unconditional quantile regressions as post-hoc analyses. We chose unconditional quantile regression because it yields more interpretable results than conditional quantile regression does. For instance, inclusion of control variables does not influence the definition of the quantile in unconditional quantile regression, while it affects the one in conditional quantile regression (Borah & Basu, 2013; Borgen, 2016; Firpo et al., 2009). We find a notable pattern in immediate effects. Our results (see Appedix B, Table 7 and Figure 1) suggest, if an inspection is announced in advance, the nursing homes whose quality level is in the middle of the distribution exhibit a significant increase in quality, while the nursing homes at the upper or lower ends of the distribution do not experience an increase. In contrast, the effects of unannounced inspections increase gradually but insignificantly along the quality level.

To address concern about issue in matching methods and procedures, we re-match treatment and control groups and consider alternative samples. First, we re-match after adjusting the operating margin. We note that TJC-accredited nursing homes have spent about $2,000 on annual accreditation fees. Our previous matching variable, operating margin, does not reflect these annual expenses. Thus, we recalculate the operating margin for the accredited nursing homes by adding $2,000 to the numerator in the operating margin formula. Next, we implement CEM and analyze the data based on the new treatment and control group. The results demonstrate the consistent result with those obtained previously. Second, in our analysis, we consider 115 nursing homes that received two inspections, namely 78 with two announced inspections, 21 that have had two unannounced inspections, and 16 nursing homes with one announced and one unannounced inspection. We verify that our results are robust when we include nursing homes that received two inspections. Third, we analyze the full sample of TJC-accredited and non-TJC-accredited nursing homes without using the matching method. We note that ex post matching approaches (e.g., coarsened exact matching and caliper-based approaches) could aggravate the imbalance between TJC-accredited and non-TJC-accredited nursing homes with respect to the variation in the unmeasured covariates (Brooks & Ohsfeldt, 2013). Our consistent results without using matching address this concern.

Finally, we conduct a series of placebo tests to investigate spurious effects of inspection visits. We first randomly assign placebo inspection visits to the treatment group, TJC-accredited nursing homes. We implement the tests for two groups of nursing homes respectively (either received an announced or unannounced inspection) with respect to qualities of all residents, long-stay residents, and short-stay residents. We obtain kernel density plots of the distribution of 2,000 placebo estimates of the inspection visits, displayed with the vertical line indicating the estimate observed in the actual data (see Appendix B, Figure 2 and Figure 3). Overall, the inspection visit estimates in the actual data are unlikely to occur by chance. To eliminate the concern in the control group, we create randomly assigned placebo visits for non-TJC-accredited nursing homes and confirm that there is no placebo effect.

3.6 Discussion

Increasingly, accreditation agencies like TJC have employed two dominant inspection strategies to ensure compliance, with the ultimate goal of improving quality performance. However, the academic literature has not distinguished between these two strategies clearly and investigated their relative benefits precisely. In this paper, we make a distinction between announced and unannounced inspections and examine their immediate and sustained effects on all-stay, long-stay and short-stay resident quality performance. The empirical analyses and supporting robustness checks show that the effects of the two strategies depend on the time duration over which the effects are aggregated (i.e. immediate or sustained) and the type of resident stay (i.e. all, short, and long-stay residents) under consideration. We discuss our results and their theoretical and practical implications for announced and unannounced inspection strategies for immediate and sustained effects, only for all-stay residents. We then discuss the nuances of these results for all, long and short-stay residents.

Our results show that as a set for all-stay residents, announced inspection strategy has significant immediate and to a lesser degree sustained quality performance effects while the effects associated with unannounced inspection strategy are not as clear. Moreover, for unannounced inspections quality performance increases steadily over time (Figure 2) for all-stay residents whereas for announced inspections, it shows a temporary spike around the time of inspection but remains relatively flat during other time periods. Using the attention-based view of a firm, we have argued that nursing home administrators have limited attention (a scarce resource), which is allocated to ensure compliance with the accreditation standards. It is reasonable to argue further that different events may require different types of attention. For instance, announced inspections may require short-term transient attention targeted to specific areas of improvement whereas unannounced inspections are likely to require more long-term, sustained attention. This logic was validated during the conversations with top managers who conceded that, unlike an announced inspection, an unannounced inspection is minimally disruptive to the general routine and employees tend to retain the same focused intensity even after the inspection is completed (JCAHO & JCR, 2006). In essence, under unannounced inspections, a nursing home is likely to achieve a steady increase in quality levels and may not experience sudden ebbs and spikes in quality levels around the inspection time. In contrast, under an announced inspection, one is likely to see detectable changes in quality levels around the inspection time, which decay to pre-inspection levels once the inspection has been conducted.

By distinguishing between announced and unannounced inspection strategies on the one hand, and the immediate and sustained quality effects on the other hand, we are able to empirically demonstrate that these two inspection modes play different roles in affecting quality performance in immediate and long terms. Findings reported in related literature do not indicate any predominant performance differences between unannounced versus announced inspections (Greenfield et al., 2007; Klerks et al., 2013), while unannounced inspections tend to be favored by the industry. Instead, studies have either shown that inspection outcomes (compliance of the standard) from announced and unannounced inspections are not statistically significantly different (Ehlers et al., 2017; Klerks et al., 2013) or the authors argue for inspection efficiency of announced inspections relative to unannounced inspections (Greenfield et al., 2012; Kim, 2015). Overall, previous studies cannot explain why unannounced inspections are prevalent in practice. However, our results show the sustained effect of unannounced inspections on quality performance that prevents degeneration in quality performance, which is not achieved by announced inspections. This result helps reveal the performance benefits of unannounced inspections. To the best of our knowledge, this is the first attempt to investigate the impacts of announced and unannounced inspections on operational performance.

In addition to all-stay residents, we also examined the above relationships for long and short-stay residents separately. While the results for long-stay residents are similar to all-stay residents under announced inspection mode (Tables 4 and 5, columns 1a and 2a), they are considerably different under unannounced mode (Tables 4 and 5, columns 1a and 2a). Our results show that an unannounced inspection strategy has a significant positive effect on both immediate and sustained quality performance for short-stay residents. These unintended positive results buttress the importance of unannounced inspection strategy for nursing home administrators.

Our findings also have important managerial implications for other stakeholders. First, our results broadly apply to other inspection agencies, whose goal is to improve quality and safety through compliance. Information reported in Table 3.1 showed that inspection strategies (announcement / unannounced) vary across regulators and industries. In practice, the decision to adopt announced, unannounced, or mixed inspection strategy requires careful consideration because unannounced inspections incur extra cost and require more effort. By demonstrating the different roles of announced and unannounced inspections, we can help regulators and inspection agencies evaluate the benefits versus the cost of these different strategies. Moreover, our findings raise an important question in supply chain context where a buyer inspects or audits a supplier: should they follow an announced or unannounced strategy?

Next, we provide crucial insight for organizations that may shun accreditation. In the healthcare industry, the accreditation rate of nursing homes is relatively low (approximately 5%), although it has increased slightly in recent years. Our findings support the view that accreditation simulates continuous improvement in quality performance for healthcare organizations (Schmaltz et al., 2011; Wagner et al., 2012b; Westphal et al., 1997). We contend that continuous improvement may stem from the accreditation that offers additional monitoring opportunity. Thus, accredited and non-accredited nursing homes should carefully reassess the cost and benefit of TJC accreditation. Furthermore, our findings have potential implications for other industries. That is, unannounced inspections may be a useful tool for maintaining consistent quality, which is critical in highly regulated industries, such as the oil and gas industry and the pipeline industry.

3.7 Limitations and Future Research Directions

This study has some limitations that could be mitigated through further research. Most of the limitations are related to the data employed in the analyses. First, we did not incorporate the outcome of inspection or accreditation in our analyses. These are important measure because unfavorable outcome could create more sustained attention after the inspection. For example, bad outcome can lead to follow-up inspections and organizations may need time to address those issues that may not be directly related to their daily operations. However, TJC does not provide specific inspection outcomes, such as the number of deficiencies or violations. Instead, TJC provides five ordinal outcomes of the accreditation on its website: accredited, accreditation with follow-up inspection, contingent accreditation, preliminary denial of accreditation, and denial of accreditation.

Although unfavorable outcome may prompt a more sustained attention to inspection on behalf of the top managers, our sample included only 11 nursing homes that received accreditation outcome other than "accredited." Thus, we cannot investigate this issue further based on the accreditation outcome. We do confirm that our results are robust after removing those 11 nursing homes.

Second, we cannot thoroughly investigate pre-inspection and post-inspection effects due to data granularity. Anecdotes from the industry suggest that the ramp-up period of an announced inspection can range from several weeks to months, while we adopted a quarter as our unit time window. Therefore, we cannot disentangle pre-inspection, inspection visit, and post-inspection effects with granularity. Future studies utilizing data incorporating each admission/discharge episode of nursing home residents (e.g., MDS data) will address this concern.

Next, we did not measure attention level of top managers before and after the two inspection regimes. Future researchers may investigate changes in attention level near the inspection time through lab or field experiments. Because attention is closely related to organizational culture, investigating culture with respect to inspection announcement can also be an interesting avenue for a future study. Interestingly, Wagner et al. (2012b) suggest that top managers in accredited nursing homes are more likely to perceive better patient safety culture in their organizations than mangers in non-accredited nursing homes. We hope that our study will spur other researchers to investigate the effect of announced and unannounced inspections in diverse contexts and across a wide range of industries using quantitative and qualitative methods.

Chapter 4
Organizational Learning, Complexity, and Safety Management Performance: Evidence from the Oil and Gas Transportation Industry

4.1 Introduction

High-hazard industries (e.g., nuclear power plants, oil production, and oil transportation) face significant threats of experiencing negative incidents, many of which are highly consequential in their impact (e.g., the Fukushima Daiichi nuclear disaster, the Deepwater Horizon oil spill, and the Kalamazoo River oil spill). These industries have experienced incidents that have resulted in significant injuries, fatalities, property loss, and environmental damage. Despite the potential for large-scale consequences, these incidents are difficult to predict and pose tremendous challenges to organizations that wish to avoid them (Leveson et al., 2009). To prevent such incidents, regulatory agencies have employed several activities, such as sharing useful information, conducting inspections, and penalizing organizations for compliance failure (Ball et al., 2017; Gray & Shimshack, 2011; Johnson, 2018; Mani & Muthulingam, 2019). In an effort to reduce accidents and improve safety, regulatory agencies have also streamlined the compliance process and created objective standards to facilitate compliance by organizations.

However, inspections are not an effective means of preventing future failure, and compliance with standards is often in conflict with learning and innovation (Ball et al., 2017; Carroll et al., 2002). To successfully pass an inspection, an organization is required to strictly adhere to regulatory standards, which in turn discourages it from experimenting with newer approaches to improve process outcomes. Additionally, inspection standards are designed to cover a broad range of organizations, are established in advance, and are changed relatively infrequently; this causes inspections to be less effective in preventing future failures, especially because organizations vary greatly from one another, and their operating environment and technology change rapidly (TRB & NASEM, 2017). Moreover, inspections are costly to regulatory agencies, which often face shortages of qualified inspectors (Congressional Research Service, 2019). To overcome these shortcomings, some regulatory agencies have proposed safety and risk management programs, which require high-hazard organizations to develop their own unique approach to identify and mitigate their specific risks within the general framework of the program. One such program, developed by the Pipeline and Hazardous Materials Safety Administration

(PHMSA) to monitor oil and gas transportation, is the Integrity Management Program (IMP). The program has attracted considerable attention because of its potential to substantially reduce regulators' monitoring burdens and costs (Kowalewski, 2013). More importantly, the Integrity Management Program constitutes a general framework, allowing organizations the opportunity to better understand their unique processes, comprehensively assess their specific hazards, and continually update their safety management procedures.

However, an extensive review of the literature reveals that there are only a handful of empirical studies in high-hazard industry settings (Haunschild & Sullivan, 2002; Madsen, 2009; Madsen & Desai, 2010; Mani & Muthulingam, 2019). Interestingly, these studies focus on learning from the failures of high-consequence, low-probability events. While the results are useful, the rarity of such events provide organizations limited opportunities to learn and innovate (Haunschild & Sullivan, 2002; Madsen & Desai, 2010; Starbuck, 2009). Moreover, few studies have examined the effectiveness of externally mandated, but internally developed risk/safety management programs, such as the Integrity Management Program (Blanco et al., 2019; Cohen & Kunreuther, 2007; Kleindorfer & Saad, 2005). We believe research is needed to better determine if an organization's participation in this type of program is universally effective, and whether there are contingencies that enhance the relationship between experience with the program and safety performance. It is important to understand these relationships, because such programs are being increasingly employed by many regulatory agencies.

We conduct our study in the context of the oil and gas pipeline industry and an operator's experience with the Integrity Management Program designed by PHMSA. Our research questions are grounded in organizational learning and complexity literature. We posit that a pipeline operator's experience with IMP has an impact on safety performance, and this relationship differs across the pipeline operator's unique structural characteristics, such as the complexity of the pipeline. We examine our questions with data from 642 pipeline operators and use several different empirical methods. Our results show that while pipeline complexity increases subsequent incident cost, experience with IMP reduces it. More interestingly, we find that complexity heightens the negative relationship between experience with IMP and future incident cost. This implies that pipeline operators with greater complexity derive greater benefits from participation in the program. Our results are robust to sample specification, sample selection bias, and endogeneity concerns (e.g., omitted variable bias). We also conduct post-hoc analyses to explore these relationships with respect to within and between pipeline operator variations using a hybrid approach. The results show that while experience with the program differs within and between pipeline operator variance, the complexity effects are universal. This implies that an operator faces higher subsequent costs from incidents as its level of complexity increases (i.e., within-operator effect), and operators with greater complexity have higher subsequent costs compared to operators with less complexity (i.e., between-operator effect). As a set, our results highlight the effectiveness of using externally mandated, but internally developed safety management programs, such as IMP, in the oil and gas transportation industry. This study contributes to the organizational learning literature, especially in high-hazard industry context. The findings are also of practical importance to both pipeline operators and federal regulators.

The rest of the paper is organized as follows. Sections 4.2 and 4.3 describe the research context, relevant literature, and hypothesis development. Section 4.4 provides the research design, empirical setting, and data, and describes the empirical strategies employed. Section 4.5 offers the results, and Section 4.6 discusses the implications and topics for future research.

4.2 Research Context

High-hazard industries include those industries involved in the production and transportation of oil and gas, nuclear, chemical manufacturing, marine transportation, air transportation, and mining sectors (TRB & NASEM, 2017). We focus on the oil and gas pipeline industry for three reasons. First, pipelines are the primary mode of oil and gas transportation, accounting for more than 70% of transportation volume in the United States. Two, oil and gas pipelines have a large footprint. In 2018, there were 2.6 million miles of oil and gas pipelines, a majority of which are distribution lines crossing every state. Finally, oil and gas pipelines carry the potential to have an immense impact on human life and property. According to the Pipeline and Hazardous Materials Safety Administration, there have been more than 700 serious incidents over the last 20 years, averaging 39 incidents, 15 fatalities, and 63 injuries per year.[27] Therefore, assessing the effectiveness of programs designed to increase safety outcomes, and identifying conditions under which they are more or less effective, is critical to evaluate.

Safety is of significant concern in the oil and pipeline industry, because it involves the transportation of hazardous materials, including crude oil, petroleum products, and liquids (e.g., propane, butane), which are highly volatile, flammable, and toxic (Kowalewski, 2013; Trench, 2003). A federal government regulatory agency, PHMSA, is tasked with developing and enforcing safety standards using inspections, among other methods.[28] However, since pipeline operators have hired away PHMSA inspectors to improve their own safety, PHMSA faces perpetual understaffing

[27] https://opsweb.phmsa.dot.gov/primis_pdm/serious_inc_trend.asp
[28] https://www.phmsa.dot.gov/about-phmsa/phmsas-mission

of well-trained inspectors (Congressional Research Service, 2019). Additionally, infrequently modified regulatory standards often fail to keep pace with the rapidly changing technological and operating environments of pipeline operators. Pipelines vary greatly from one another in terms of dimensions (e.g., narrow, medium, wide), materials used in their construction, and geographic dispersion (e.g., within state or across state boundaries), exacerbating the mismatch between operating procedures and regulatory standards, which are designed to apply to many different types of organizations. Such standards are ill-suited to accurately assess risk and prevent subsequent incidents. They also hinder regulators' ability to establish a magic bullet (e.g., micro-level regulatory standards). Previous studies have also noted that inspections and penalties are poor deterrents in reducing incidents (Stafford, 2014, 2017).

To overcome these problems, PHMSA introduced the Integrity Management Program in 2001. IMP is a proactive program that calls for pipeline operators to establish their own customized internal safety management programs to identify and mitigate risks (DeWolf, 2003; TRB & NASEM, 2017). Specifically, operators of hazardous liquid pipelines have been required to "identify, prioritize, assess, evaluate, repair and validate the integrity of hazardous liquid pipelines that could affect High Consequence Areas (HCAs)." HCAs represent densely populated area and areas with drinking water and ecologically sensitive resources (Kowalewski, 2013, 49 C.F.R. §195, 2006). Under this program, PHMSA assigns responsibility to pipeline operators for managing their own risks in a systematic way because they are in a better position to understand the sources of risk in their system and take necessary action (Kowalewski, 2013; TRB & NASEM, 2017). In this way, operators can prioritize resources to HCAs and reduce potential damage from incidents. IMP also affords operators the flexibility to choose their own risk reduction technologies and practices within the general guidelines. This allows them to customize the program to suit their own unique needs and characteristics. Overall, IMP is a safety management program that helps operators learn about their processes.

4.3 Literature Review, Theoretical Grounding, and Hypotheses

Our research is informed by two streams of literature: structural complexity and organizational learning, both of which are mature in nature and have extensive functional spans. In reviewing them, we focused our efforts on their relevance to our chosen context, high-hazard industries.

Complexity

Complexity is frequently associated with “a large number of parts that interact in a non-simple way” (Simon, 1962). Other researchers have focused on the heterogeneous aspect of complexity, defining it as “the degree of heterogeneity in the range of activities” that are relevant to an organization’s operations. Heterogeneity, here, is often described as “the degree of similarity between constituent elements” (Aldrich, 1979; Child, 1972; Dess & Beard, 1984). Our review of the operations management literature reveals that complexity has been measured using many different units of analysis (Table 4.1). These include task complexity (Argote et al., 1995; Avgerinos & Gokpinar, 2017), product and process complexity (Novak & Eppinger, 2001; Vickery et al., 2016; Wolf, 2001), project complexity (Peng et al., 2014), supply chain complexity (Bode & Wagner, 2015; Bozarth et al., 2009; Choi & Krause, 2006), and environmental (market) complexity (Azadegan et al., 2013; Wiengarten et al., 2017). The literature is conclusive in terms of its impact on performance: researchers agree that complexity diminishes performance, and the relationship is true across different measures of complexity and performance. For instance, high task complexity is associated with lower productivity (Argote et al., 1995), and supply chain complexity (e.g., upstream, manufacturing, and downstream) lowers plant performance (Bozarth et al., 2009). Additionally, complexity in the process and supply chain leads to more frequent disruptions and higher incident rates (Bode & Wagner, 2015; Wolf, 2001). The basic argument underlying the causes of the negative impact is that because complexity entails numerous components with multiple interactions, it is difficult to comprehend all of the information necessary to process it. This makes it impossible for an individual to perceive inherent risks and execute appropriate tasks, resulting in inferior performance (Azadegan et al. 2013, Child 1972, Dess and Beard 1984, Peng et al. 2014, Vickery et al. 2016, Wood 1986).

Complexity is inherent in high-hazard activities and organizations (TRB & NASEM, 2017). Specifically, Perrow (1999) proposes that complexity in a system will increase risk where the interactions in the system “cannot be thoroughly planned, understood, predicted, or guarded against” (Leveson et al., 2009). Oil and gas pipelines are inherently complex because they consist of a variety of physical components (e.g., pipes, valves, break tanks, and other connected structures) and vary in terms of construction methods, physical properties, design configurations, and operating conditions. Most of the pipes are buried and are therefore invisible, making it challenging to monitor, maintain, and repair them. Operators may also face negative repercussions when they point out problems or potential risks in the pipelines, reducing the likelihood of identifying problems (Leveson et al., 2009; TRB & NASEM, 2017). Pipelines are operated by humans who are prone to mistakes (Kowalewski, 2013). The above logic suggests that an increase in pipeline complexity has a negative impact on performance and results in a higher incident risk.

H1. Complexity is positively associated with subsequent incidents.

Table 4.1 Overview of Role Complexity on Operational Performance

Authors (year)	Industry Sector (Unit of analysis)	Research Focus	Type and source of complexity	Dependent Variables
Argote et al. (1995)	Experiment (240 subjects)	Effects of turnover and task complexity on group performance – Complexity reduces productivity, and this becomes more apparent through repetitive tasks over time (learning). – Complexity dampens the effect of turnover on group productivity.	Task complexity (required steps in production)	Productivity (production volume)
Avgerinos & Gokpinar (2017)	Healthcare (6,206 surgeries)	Effect of task complexity and task familiarity on team productivity – Complexity enhances the positive relationship between team familiarity and team productivity	Task complexity (patient condition)	Productivity (surgery duration)
Azadegan et al. (2013)	Manufacturing (126 firms)	Effect of environmental complexity and environmental dynamism on lean practices – A positive effect of complexity on performance – Lean practices mitigate the negative effect of environmental complexity on performance	Environmental (market) complexity (-industry concentration index)	Performance (gross margin)
Bode & Wagner (2015)	Manufacturing (396 buying firms)	Effect of supply chain complexity on the frequency of supply chain disruptions – All supply chain complexity measures leads to more disruptions.	SC complexity (horizontal, vertical, spatial)	SC disruptions
Bozarth et al. (2009)	Manufacturing (209 plants)	Effect of supply chain (SC) complexity on plant performance – All complexity constructs adversely influence plant performance.	SC complexity (upstream, manufacturing, downstream, dynamic)	Operational performance (e.g., cost)
Liu (2015)	Information system (128 projects)	Contingent effect of complexity risk on the relationship between control and project performance – A negative effect of complexity on performance – Moderating effects of complexity risk are different across control modes (positive: behavior/self-control, negative: outcome/clan control).	Complexity risk (perceived technology immaturity, uncertainty)	Perceived project performance
Lo et al. (2014)	Manufacturing (211 firms)	Relationship between a safety certification and operational performance – Complexity heightens the effect of certification on operational performance.	Complexity (R&D and labor intensities)	Operational performance (profitability, productivity, safety)
Peng et al. (2014)	Manufacturing (212 NPD projects)	Relationship between IT tools and NPD collaboration – Project complexity dimensions moderate this relationship: positively (product size), negatively (project novelty, task interdependence).	Project complexity (product size, project novelty, task interdependence)	Perceived NPD collaboration
Vickery et al. (2016)	Manufacturing (112 firms)	Relationship between product modularity, process modularity, and NPD performance – An insignificant main effect of complexity on NPD performance. – Complexity negatively (positively) moderates the effect of product (process) modularity on NPD performance.	Product/process complexity	Perceived NPD performance
Wiengarten et al. (2017)	Manufacturing (3,945 firms)	Effects of operational and financial slacks on occupational safety – Complexity positively moderates the relationships between coupling and safety violations.	Environmental (market) complexity (-industry concentration index)	Safety (violations)
Wolf (2001)	Petrochemical (36 refineries)	Relationship between complexity, coupling, and safety performance – Chemical release accident rates is related to the refinery complexity, but occupational safety incident rate is not.	Process complexity (# of possible states in a process system)	Incident rate

Organizational Learning in High-hazard Industries

Organizational learning refers to "a process of detecting and correcting error," and experience is fundamental to the acquisition of knowledge and learning (Argyris, 1977; Huber, 1991). Organizational learning has been studied in a variety of fields, demonstrating that organizations improve their performance through acquired knowledge and learning (Argote et al., 2003; Huber, 1991; Madsen, 2009; Mani & Muthulingam, 2019). However, there are two notable trends in the learning literature. First, while early studies have focused on the benefits of learning in terms of cost and productivity, more recent studies concern improvements in other performance areas, such as quality and safety (Argote et al., 1995; Ball et al., 2017; Baum & Dahlin, 2007; Haunschild & Sullivan, 2002; KC et al., 2013; Madsen, 2009; Mani & Muthulingam, 2019). Next, recent studies have disaggregated learning experience. For instance, these studies decompose whether organizations learn from the following: their own experience (direct learning) or the experience of others (indirect learning); experience with success or failure; experience at a specific site; and voluntary or involuntary experience (Ball et al., 2017; Baum & Dahlin, 2007; Haunschild & Rhee, 2004; Haunschild & Sullivan, 2002; Huber, 1991; Levitt & March, 1988; Madsen, 2009; Madsen & Desai, 2010; Mani & Muthulingam, 2019).

However, organizational learning in high-hazard organizations is more idiosyncratic. Learning from failures in the form of high-consequence, low-probability events is often problematic, because these rare events provide only a few opportunities to learn and little motivation to prepare for future risk (Starbuck, 2009). We also lack relevant empirical studies, because existing literature has focused on qualitative studies (Baum & Dahlin, 2007; Carroll et al., 2002; Labib et al., 2019). However, we review notable empirical studies in this domain (e.g., pharmaceutical, airline, orbital vehicle, coal mining, oil production) listed in Table 4.2. The review presents important findings which indicate that high-hazard organizations learn more effectively under certain conditions. Specifically, they learn more from their own failures than success (Madsen & Desai, 2010; Mani & Muthulingam, 2019); from complex causes than simple causes (Haunschild & Sullivan, 2002); from major failure than minor failure (Madsen, 2009); from firm-specific experience than other experience (Huckman & Pisano, 2006); and from voluntary experience than involuntary experience (Haunschild & Rhee, 2004).

Among various learning contexts in high-hazard industries, we study organizational learning from IMP, an internally developed safety management program, because this program facilitates higher-level learning more so than other contexts (e.g., failure, inspection). For instance, when organizations attempt to learn from failures, they often learn superficially (Reason, 1997). Further, failure investigation and routine inspections are external events that create defensive responses from organizations (Haunschild & Rhee, 2004; Marcus, 1988). Specifically, mechanical compliance will induce similar future errors, because unwilling responses will not create retainable knowledge that permanently influences organizational routines (Haunschild & Rhee, 2004; Perrow, 1999). In contrast, when organizations internally develop a safety framework and practices, such as IMP, those efforts penetrate to organizational routines (Haunschild & Rhee, 2004; Marcus & Nichols, 1999). While PHMSA allows flexibility in implementing IMP, internally developed IMP per a pipeline operator should contain the elements of identifying, prioritizing, assessing, evaluating, repairing, and validating the integrity of hazardous liquid pipelines that could affect high consequence areas (See detail steps in Figure 1) (Kowalewski, 2013; 49 C.F.R. §195, 2006). Through these steps in IMP, operators can develop a sufficient understanding of the appropriate preventive and mitigative actions and integrate the program into their daily activities (DeWolf, 2003).

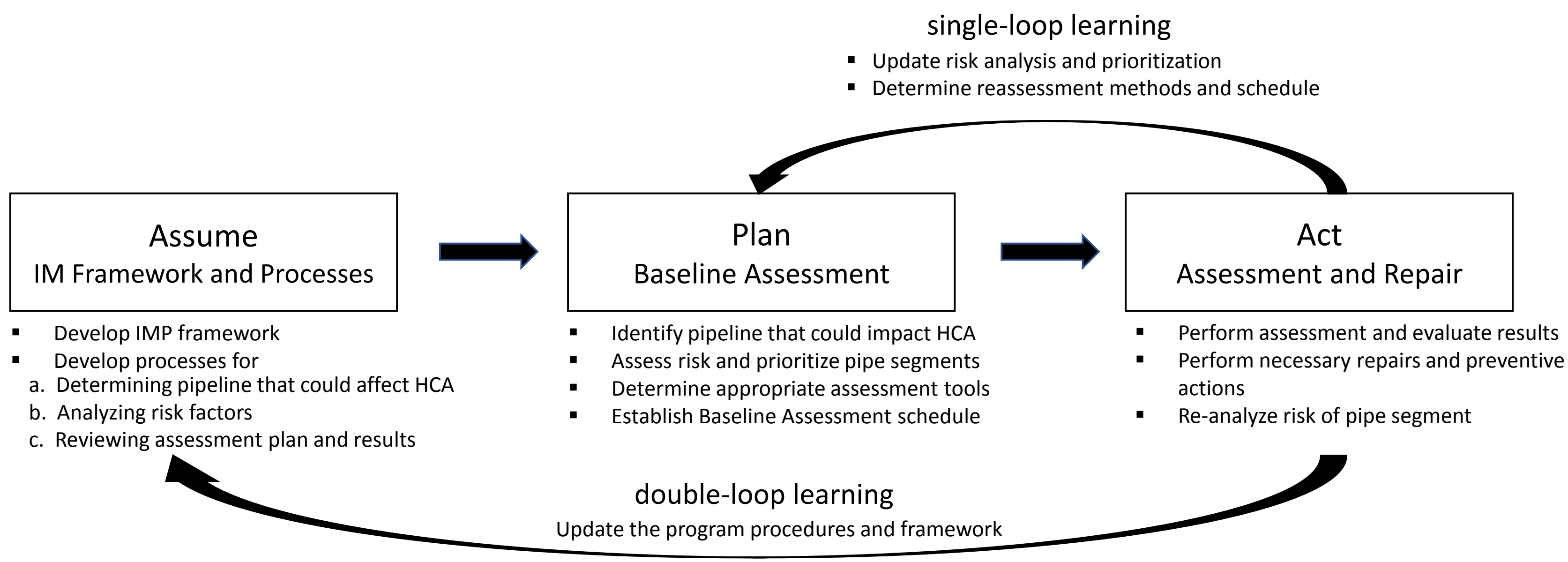
Consistent with organizational learning theory, pipeline operators engage in single-loop and double-loop learning through IMP. Single-loop learning denotes that organizations adjust actions based on feedback to fill the gap between the desired goal and actual achievement (Argyris, 1977; Carroll et al., 2002). In our context, described in Figure 4.1, operators update risk analysis, establish reassessment tools, and schedule future plans (baseline assessment plans) based on their assessment and necessary repair (action), a process which corresponds to single-loop learning. In contrast, double-loop learning refers to question and challenge underlying assumptions, values, and the appropriateness of goals (Argyris, 1977; Carroll et al., 2002). IMP requires operators to continually update program procedures and frameworks (assumptions) to reflect their previous assessment and repair (act), which generates double-loop learning (Figure 4.1). Overall, as operators gain experience with IMPs, their IMPs have evolved and improved to better mitigate risk and enhance safety. Therefore, we argue that as operators gain learning experience from the Integrity Management Program, their safety performance increases, lowering incident risk.

H2. Organizational learning is negatively associated with subsequent risk.

Table 4.2 Overview of Organizational Learning in Safety and High-hazard Industries

Authors (year)	Industry Sector (Unit of analysis)	Research Focus and Findings	Independent Variables (Type of learning)	Dependent Variables
Ball et al. (2017)	Pharmaceutical (2,244 plants)	Effect of inspectors' learning on future recalls – Site-specific inspection experience of inspectors lead to more subsequent recalls.	Learning (experience in inspection – direct, site-specific or other)	Recall frequency
Baum & Dahlin (2007)	Freight rail (189 railroad-year)	Effect of learning and aspiration-performance feedback on – Own operating experience and others' accident experiences reduce subsequent accident consequence. – Aspiration-performance feedback (e.g., gap from historical accident rate) strengthens the learning effect on accident consequence.	Learning (operating and accident experience – direct, indirect, success, failure)	Accident consequence (cost per operating mile)
Haunschild & Rhee (2004)	Automotive (47 automakers)	Effect of learning with automobile recalls on future recalls – Production and voluntary recall experiences decrease subsequent involuntary recalls.	Learning (production and (in)voluntary recall experience – direct, indirect, success, failure)	Recall frequency
Haunschild & Sullivan (2002)	Airline (310 airlines)	Effect of learning from heterogeneity on accident rate heterogeneity is generally better for learning, as prior heterogeneity in the causes of errors decreases subsequent accident rates. – Own accident experience reduces subsequent accident rate. – Airlines learn more from accidents with heterogeneous (complex) causes than one with a homogeneous cause.	Learning (experience in accidents with (homo-)heterogeneous causes – direct, failure)	Accident rate (accidents per 100k departures)
Huckman & Pisano (2006)	Healthcare (203 surgeons)	Effect of freelancing (site-specific or other experience) on mortality – Surgeons learn from their experiences with surgeries at local hospital (site-specific experience) not at other hospitals.	Learning (experience in surgeries – direct, site-specific or other)	Mortality
KC et al. (2013)	Healthcare (71 surgeons)	Effect of individual learning from own and others' success and failure on mortality – Surgeons learn more from their own successes than failures. – Surgeons learn more from the failures of others than from successes.	Learning (experience in surgeries – direct, indirect, success, failure)	Mortality
Madsen & Desai (2010)	Orbital vehicle (36 organizations)	Effects of learning from success and failure in orbital launch – Organizations learn more effectively from failures than successes – The effect of learning from failures depreciates more slowly than from successes.	Learning (experience with success failure – direct, indirect, success, failure)	Failed or succeeded launch
Madsen (2009)	Coal mining (20,864 mines)	Effects of learning from minor accidents and disasters experience – Own experience and the experience of others prevent future disasters. – The effect of learning from minor accidents depreciates more rapidly than from disasters.	Learning (experience with minor accident and disasters – direct, indirect, failure)	Accident frequency (disasters)
Mani & Muthulingam (2018)	Oil production (13,606 wells)	Effect of learning from own and others' inspection experience on environmental performance – A focal unit learns from inspection of other units when inspections detect violations (failure).	Learning (experience in inspection – direct, indirect, success, failure)	Environmental performance (inspection violations)

Figure 4.1 Learning in Integrity Management Program



Complexity in Learning

Researchers in organizational learning literature note the difficulty in learning from experience. Organizational environments are often complex, which renders their experience ambiguous (Levinthal & March, 1993; Levitt & March, 1988). However, previous studies do not provide a coherent theory and accompanying evidence to indicate whether complexity facilitates or hinders organizational learning. Instead, complexity has been considered a double-edged sword in learning (Rijpma, 1997). On the one hand, scholars lay out two argument to prove that complexity heightens the effect of organizational learning. First, complexity fosters learning because it requires organizations to broaden their perspectives and create greater desire and need to learn (Carroll et al., 2002; Cohen & Levinthal, 1990; Haunschild & Sullivan, 2002; Rijpma, 1997). Second, complexity offers greater learning opportunities because it leads to a deeper understanding and analysis of the problem. Some studies argue that the effects of learning with experience (e.g., airline failures, safety certification) are stronger when the learning environment is complex (Haunschild & Sullivan, 2002; Schilling et al., 2003; Vickery et al., 2016). On the other hand, learning under complex conditions is difficult and ambiguous, and organizations sometimes struggle to comprehend knowledge derived from complex situations (Baum & Dahlin, 2007; Carroll et al., 2002; Perrow, 1999; Rijpma, 1997). When organizations operate complex systems or encounter complex situations, they have difficulties identifying the best solution and sometimes view systematic failures as random (Haunschild & Sullivan, 2002; Pisano et al., 2001). However, relevant empirical evidence is scarce (Argote et al., 1995). Taken together, and consistent with more concrete evidence that supports the beneficial role of complexity, we argue that pipeline operators with highly complex system will be exposed to greater learning opportunities to broaden their perspectives and practices. Thus, the effect of organizational learning is greater for pipeline operators with high complexity rather than low complexity.

H3. The effect of organizational learning on subsequent risk is greater under high complexity rather than low complexity.

4.4 Research Design and Data

Sample and Data Collection

We investigate the primary effects of complexity and organizational learning and the moderation effect of complexity among hazardous material (hazmat) pipeline operators in the United States. We chose hazmat pipeline operators because of their immense consequences to human health, property, and the environment, as they operate about 200,000 miles of pipeline (Kowalewski, 2013; Stafford, 2014). To test our hypotheses, we use data from the Pipeline and Hazardous Materials Safety Administration (PHMSA), and ORBIS, a private company database of the Bureau van Dijk Electronic Publishing firm that compiles administrative data from operators. The primary data comes from PHMSA reports, which have been completed by hazmat pipeline operators. Reporting is mandatory for operators, because PHMSA will impose a civil penalty for failure to report pursuant to the Code of Federal Regulations (49 C.F.R. §195, 2006). We first obtained pipeline operators' characteristics and their activities in the Integrity Management Program from operator-level annual reports. Second, we retrieved incident information (e.g., date, operator, consequence of incident) from incident reports. We then merged these data using the unique identification numbers of the operators assigned by PHMSA. Third, we complement other characteristics of pipeline operators (e.g., ownership of operator and parent company) from ORBIS by matching the names and addresses of pipeline operators between PHMSA reports and ORBIS database. We then excluded those pipeline operators when characteristics of the incident (e.g., total pipeline mileage, nominal size) were missing. This process resulted in the final sample, an unbalanced panel of 642 hazmat pipeline operators, consisting of 4,696 pipeline operator-year observations between 2004 and 2017. We describe corresponding variables and provide summary statistics in Table 4.3.

We use data range from 2004 to 2017 because PHMSA carried out major revision to the reports, and this revision affected operators' submissions from 2004 forward.[29] The unit of analysis is pipeline operator, because annual reports contain pipeline characteristic information at the pipeline operator level, and we consider a pipeline operator a learning unit. This consideration is appropriate because a pipeline operator operates an entire network system, including but not limited to line pipe, valves, pumping units, and other equipment connected to the line pipe (49 C.F.R. §195, 2006).

[29] We note that the initial IMP activities in 2002 and 2003 were recorded in the report of year 2004, but additional pipeline characteristics have been reported since 2004.

Table 4.3 Description of Variables and Summary Statistics

Variable	Description (n = 642, *Observations* = 4,696)	Mean	SD	Min	Max
Safety Performance Measures					
Incident CostPerMile $_{t+1}$	Log of (incident cost per pipeline miles) on next year	1.266	2.632	0	14.024
Independent Variables					
Complexity	Degree of complexity in a pipeline system, measured by variability from installed years across segments	0.323	0.306	0	0.883
BaseAssess Indicator	1 if an operator conducts Baseline Assessment	0.397	0.489	0	1
BaseAssesst Experience	Log of pipeline mileages that an operator conducts baseline assessment	1.343	1.995	0	8.380
Yrs from Last BaseAssess	Years passed since the last Baseline Assessment	1.100	1.989	0	14
Controls					
Ownership	Operator is a public company (Yes = 46.4%)				
HCA Ratio	Pipeline miles in High Consequence Area over total miles	0.523	0.395	0	1.000
Pipeline Diameter Large	Pipeline miles with diameter ≥ 18" over total miles	0.098	0.252	0	1.000
Pipeline Diameter Medium	Pipeline miles with diameter between 10" and 18" over total miles	0.238	0.338	0	1.000
Average Age	Pipeline average age from installation year	36.664	18.897	0	101.000

Table 4.4 Correlation Matrix

	ICC	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) *Incident CostPerMile* $_{t+1}$	0.483	1									
(2) *Complexity*	0.890	0.346**	1								
(3) *BaseAssess Indicator*	0.302	0.258**	0.375**	1							
(4) *BaseAssess Experience*	0.521	0.355**	0.427**	0.829**	1						
(5) *Yrs from Last BaseAssess*	0.452	-0.111**	-0.139**	-0.449**	-0.372**	1					
(6) *Ownership*	0.838	0.048**	0.143**	0.069**	0.063**	-0.009	1				
(7) *HCA Ratio*	0.802	-0.007	0.062***	0.091***	-0.001	0.069***	-0.016	1			
(8) *Pipeline Diameter Large*	0.975	0.170***	-0.046**	0.063***	0.125***	-0.005	-0.004	0.010	1		
(9) *Pipeline Diameter Medium*	0.934	0.079***	0.008	0.045**	0.055***	0.047**	0.037*	0.064***	-0.130***	1	
(10) *Average Age*	0.824	0.205***	0.310***	0.185***	0.212***	0.005	0.021	0.166***	0.022	0.002	1

Intraclass Correlation Coefficient (ICC) denotes between variation over total variation (sum of between and within variation). **p < 0.01, *p <0.05.

Dependent Variables

Safety Performance: Incident Cost Per Mile. We consider a subsequent incident a key construct in the safety management performance of pipeline operators, because it reflects the outcome of the managerial effort of the operators (Sosa & Alvarez-Ramirez, 2009; Stafford, 2014). In the pipeline industry, federal regulations refer to an incident as an unintentional release of a hazardous material (49 C.F.R §171.16, 2011). While there are two types of measures with respect to safety, incident consequences (e.g., incident cost, spilled volume) and incident frequency (e.g., incident rate), we use a consequence measure of cost, which has been employed in previous studies (Baum & Dahlin, 2007; DeWolf, 2003; Leveson et al., 2009). We focus on the cost measure because it can comprehensively capture the influences of incidents on operators, the public, and the environment, while a frequency measure cannot precisely assess these impacts. Further, the frequency measure is imperfect because a comparison of incident frequency only makes sense if the consequences stemming from the incident are similar (Leveson et al., 2009). Further, the cost measure also outperforms the other consequence measure, spilled volume, for several reasons. First, the volume measure does not recognize the impacts of different commodities when the volume spilled is the same. However, the total cost of releasing highly volatile liquids (e.g., propane, butane) is radically different than a release of crude oil. Next, the volume measure does not reflect whether hazmat is released in high consequence areas (e.g., densely populated area) or low consequence areas. Instead, the cost measure quantifies the overall consequence through costs. To create the cost measure, we first compute the total cost (indexed to 2018 dollars) of pipeline incidents over pipeline miles per operator from PHMSA reports, then take a logarithmic form. Specifically, we take together the costs of an operator's lost commodity, property damage and repairs, emergency response, and environmental remediation, and the cost of third-party property damage (e.g., public property, non-operator private property) from the incidents. To account for impacts on subsequent risk, we consider the incidents cost per mile at t+1 year.

Independent Variables

Complexity. Organizational theory has conceded that complexity is a structural variable which portrays organizations (Anderson, 1999). As Bar-Yam (1997) considers complexity as the amount of information needed to describe a system, researchers have suggested that homogeneity-heterogeneity and concentration-dispersion describe complexity (Dess & Beard, 1984; Haunschild & Sullivan, 2002). Among various factors that generate heterogeneity in pipeline operator systems, we regard heterogeneity from the year of pipeline installed (age) as a key measure to construct structural complexity. This is because the risk factors from pipeline characteristics are interdependent, and age is a clear and logical indicator of physical properties (e.g., materials, construction methods) and the way to assess pipeline condition (e.g., hydrostatic test) (Kiefner & Trench, 2001). For instance, pipelines from the early 1900s still contain iron; pipelines from the 1920s were constructed with electric-resistance welding; pipelines from the 1940s were built with cathodic protection to prevent corrosion; pipelines from the late 1960s were coupled with high frequency electric-resistance welding; and older pipelines do not conform to in-line cleaning and inspection (Kiefner & Trench, 2001; TRB & NASEM, 2017). The heterogeneity in pipeline age reflects a complexity in pipeline operations that requires different monitoring, maintenance, and repair practices.

To create a structural complexity measure from age heterogeneity, we use information on pipeline mileage by installed decade per operator (i.e., mileage of pipeline installed from 1920-1929) contained in PHMSA annual reports. We measure the complexity by computing a Herfindahl-Hirschman Index (HHI) from pipeline mileage by installed decade and then subtracting the HHI from 1.

Learning. To capture organizational learning through the Integrity Management Program, we focus on the baseline assessment for two reasons. First, this is a primary step that builds a technical basis for determining risk factors used in scheduling and selecting integrity assessment methods, and that conducts scheduled assessments to understand the condition of the pipeline (TRB & NASEM, 2017). Second, this is a step, which has been observed by PHMSA and recorded in the annual reports. Therefore, it is an appropriate proxy for overall learning from IMP. However, we note a unique setting in the implementation of baseline assessment under federal regulations (49 C.F.R. §195, 2006). The regulations require operators to maintain intervals between assessments of less than 68 months, although operators may lengthen the interval and defer the next assessment with engineering-based justification. In that sense, we construct two main variables with one control variable. First, we create a dichotomous variable of *BaseAssess Indicator* that indicates whether a pipeline operator conducts a baseline assessment in that year. This represents whether pipeline operators encounter opportunity to learn by doing (baseline assessment) in that year. Second, we create a volume measure of *BaseAssess Experience* by computing logged pipeline miles for which an operator conducts a baseline assessment. We focus on experience at the current time frame instead of cumulative experience, because operators may have difficulty retrieving older experiences, and experiences may not be preserved when the level of technology keeps changing (Levitt & March, 1988). Finally, we create a control variable, *Years from Last BaseAssess*, because an operator will conduct the next baseline assessment after a certain number of years which probably will not exceed 68 months, as noted in the regulation.

Control Variables

We incorporate key control variables from previous literature and federal law about the transportation of hazardous liquids by pipeline (DeWolf, 2003; Stafford, 2014, 2017; 49 C.F.R. §195, 2006). This law includes risk factors delineated by the Department of Transportation, the National Transportation Safety Board, the Environmental Protection Agency, and the Technical Hazardous Liquid Pipeline Safety Standards Committee. We also include other risk factors, which might influence risk management efforts and outcomes.

Ownership. We control for the ownership of operators because we expect that public companies will be more concerned about incidents that may result in a potential drop in their stock price. Since many pipeline operators are private companies, we comprehensively investigate the ownership of the parent company. We collect this information from Orbis and code an ownership dummy variable as one if either an operator or its parent company is a public company and otherwise zero.

HCA Ratio. There are areas in which incidents will have a more significant impact, even with the same amount of hazardous material release, because such areas are densely populated or environmentally sensitive. Specifically, federal law defines a high consequence area (HCA) as a highly populated area, an unusually sensitive area, or a commercially navigable waterway (49 C.F.R. §195, 2006). HCA is an important risk indicator, as noted in the regulation, and it requires greater attention and protection from incidents (Kowalewski, 2013; 49 C.F.R. §195, 2006). To control for the *HCA Ratio*, we create a measure by computing pipeline mileage at HCA over total pipeline miles per operator.

Pipeline Diameter. PHMSA suggests that an operator consider pipeline diameter size when it establishes baseline assessment and IMP (49 C.F.R. §195, 2006). While the range of pipeline diameter usually spans from 8 to 42 inches, the federal regulations consider the pipe segments with a diameter greater than or equal to 18inches as high risk, a diameter between 10 inches and 18inches as moderate risk, and a diameter of less than 10 inches as low risk. To control for risk from pipeline diameter size, we collect pipeline mileage by diameter size per operator (e.g., mileage of pipeline whose nominal size is less than 10 inches) from the annual reports. We then create two variables, each of which computes the pipeline mileage rate whose diameter is within a certain range (e.g., greater than 18 inches, between 10 inches and 18 inches) over the total pipeline mileage of an operator, respectively.

Average Age. PHMSA requires a pipeline operator to consider pipeline age when it establishes an internal assessment of risks and plans (49 CFR §195.452, 2010). While federal law indicates that a pipeline segment older than 25 years has a high risk, previous literature has argued that pipelines installed before 1930 have consistently more incidents because they have experienced more external corrosion (Kiefner & Trench, 2001; Kowalewski, 2013). However, to incorporate pipeline age comprehensively, we calculate the weighted average using decade pipeline installed and its mileage value (i.e. mileage of pipeline installed during 1920-1929) in the annual reports. We then subtract the calculated average from the current year of observation.

Empirical Strategy

Our analyses investigate three research questions: whether the relationship between complexity and subsequent risk is positive; whether the relationship between organizational learning from the internally developed program (IMP) and subsequent risk is negative; and when organizational learning becomes more effective (e.g., level of complexity). We establish ordinary least squares regression models to test the main effects of complexity ($Complexity_{it}$) and baseline assessment experience (*BaseAssess Indicator*, $BaseAssess\ Experience_{it}$). We then investigate the moderating role of complexity using subgroup (segmented regression) analysis by splitting samples into two groups depending on the level of complexity. We employ subgroup analyses in lieu of interaction variables because interaction effects in fixed-effect models are likely to be biased (Giesselmann & Schmidt-Catran, 2018; Shaver, 2019). Specifically, interaction terms in fixed-effect models do not purely capture within-group variation. Instead, the estimates of interaction effects will exhibit both within-group and between-group variations. To avoid ambiguous results and interpretation, we employ subgroup analysis, as suggested by Shaver (2019). For all regression models, we include time-varying characteristics of pipeline operators ($Controls_{it}$) and fixed effects of pipeline operator and year. The following equation describes our model, where $Incident\ CostPerMile_{t+1}$ is the logged amount of incident cost per pipeline mileage of an operator i at year t+1, while other variables are constructed from an operator i at year t. We use standard errors clustered in operator-level.

$$Incident\ CostPerMile_{it+1} = \beta_1 Complexity_{it} + \beta_2 BaseAssess\ Indicator_{it} + \beta_3 BaseAssess\ Experience_{it} + \boldsymbol{\gamma} Controls_{it} + \boldsymbol{\delta} Operator\ Fixed\ Effect_i + \boldsymbol{\tau} Year\ Fixed\ Effect_t$$

4.5 Results

A correlation matrix table in Table 4.4 suggests that both complexity and baseline assessment variables are positively associated with the dependent variable: *Complexity* and *Incident CostPerMile* ($\rho = 0.346$, $p < 0.01$), *BaseAssess Indicator* and *Incident CostPerMile* ($\rho = 0.258$, $p < 0.01$), *BaseAssess Experience* and *Incident CostPerMile* ($\rho = 0.355$, $p < 0.01$). We observe that *Complexity* and *BaseAssess* variables are positively associated: *Complexity* and *BaseAssess Indicator* ($\rho = 0.375$, $p < 0.01$), *Complexity* and *BaseAssess Experience* ($\rho = 0.427$, $p < 0.01$).

To test our hypotheses, we first use a full sample: 642 pipeline operators with 4,696 operator-year observations. We then consider a subsample, which excludes pipeline operators that never experienced any incidents or that never conducted baseline assessment activities during the research period. This consideration reduces the number of pipeline operators from 642 to 234, consisting of 2,369 operator-year observations (see Appendix C, Table 1). We report primary analyses from the testing of H1 and H2 in Table 4.5 and from the testing of H3 in Table 4.6, for both the full sample and the subsample.

We first describe the main effects of complexity and the baseline assessment on incident cost in Table 4.5. Models 1, 2 and 3 denote the results from the full sample, and Models 4, 5, and 6 represent the results from the subsample. We first include complexity with control variables (Model 1 – full sample; Model 4 – subsample) and baseline assessment variables with controls (Model 2 – full sample; Model 5 – subsample), respectively. We then consider both the effects of complexity and a baseline assessment in a stepwise manner (Model 3 – full sample; Model 6 – subsample). Our results are consistent across the full sample and the subsample. We find that *Complexity* is positively associated with subsequent incident cost ($b_1 = 1.042$, $p < 0.01$, Model 1; $b_1 = 2.085$, $p < 0.01$, Model 4), supporting H1, which theorized a positive effect. Next, while the coefficient of *BaseAssess Indicator* is positive and statistically significant ($b_2 = 0.330$, $p < 0.01$, Model 2; $b_2 = 0.592$, $p < 0.01$, Model 5), the estimate of *BaseAssess Experience* is negative and significant, supporting H2 ($b_3 = -0.107$, $p < 0.05$, Model 2; $b_3 = -0.159$, $p < 0.05$, Model 5). These effects of complexity and baseline assessment remain significant in Models 3 and 6.

Next, we evaluate the moderation effect of complexity on the relationship between baseline assessment and subsequent incident cost (H3). We report subgroup analyses in Table 4.6: Models 1, 2 and 3 denote the results from the full sample, and Models 4, 5, and 6 represent the results from the subsample. We include Models 1 and 4 in Table 4.6, which correspond to Models 3 and 6 in Table 4.5, to render the main estimates comparable to the estimates from subgroup analyses. We lay out the results of subgroup analyses, when pipeline structural complexity is high (Model 2 – Full sample; Model 5 – subsample) and low (Model 3 – Full sample; Model 6 – subsample). We show consistent results across the full sample and the subsample. When the complexity level is high (Models 2 and 5), we observe a non-significant positive association between *Complexity* and subsequent incident cost ($b_1 = 0.733$, $p > 0.\ 1$, Model 2; $b_1 = 1.086$, $p > 0.\ 1$, Model 5; H1 not supported). The results also suggest no significant association between *BaseAssess Indicator* and subsequent incident cost ($b_2 = 0.304$, $p > 0.10$, Model 2; $b_2 = 0.113$, $p > 0.10$, Model 5). However, *BaseAssess Experience* and subsequent incident cost show a negative and significant association ($b_3 = -0.141$, $p < 0.05$. Model 2; $b_3 = -0.169$, $p < 0.05$. Model 5; H2 supported). In contrast, when we consider operators with low complexity (Models 3 and 6), we find a strong and positive association between *Complexity* and subsequent incident cost ($b_1 = 1.323$, $p < 0.01$, Model 3; $b_1 = 2.558$, $p < 0.01$, Model 6; H1 supported). We observe a positive association between *BaseAssess Indicator* and subsequent incident costs without significance in the full sample ($b_2 = 0.221$, $p > 0.10$, Model 3), and with significance in the subsample ($b_2 = 0.809$, $p < 0.05$, Model 6). The results further suggest an insignificant negative association between *BaseAssess Experience* and subsequent incident costs ($b_3 = -0.041$, $p > 0.10$, Model 3; $b_3 = -0.162$, $p > 0.10$, Model 6). To investigate the moderation effects, we compare coefficients across operator groups with high and low complexity, following the procedure introduced by Clogg et al. (1995). While the differences in main effects between high and low complexity groups are mostly insignificant, *BaseAssess Experience* in high and low complexity groups is significantly different ($\Delta b_2 = -0.696$, $p < 0.10$, Models 5 and 6). Interestingly, the main effects of complexity and the baseline assessment variables are not universally significant between the two groups. Specifically, the effect of complexity is insignificant in high complexity groups but significant in low complexity groups for both the full sample and the subsample. In contrast, the effect of *BaseAssess Experience* is significant in high complexity groups but insignificant in low complexity groups.

Table 4.5 Regression Analyses Testing Complexity (H1) and Learning (H2)

Sample	Full-Sample			Sub-Sample with no incidents; no IMP		
DV: *Incident CostPerMile*$_{t+1}$	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Complexity	1.042**		1.034**	2.085**		2.130**
	(0.335)		(0.332)	(0.595)		(0.584)
BaseAssess Indicator		0.330*	0.320*		0.592*	0.567*
		(0.155)	(0.155)		(0.266)	(0.265)
BaseAssess Experience		-0.107*	-0.110*		-0.159*	-0.170**
		(0.047)	(0.047)		(0.062)	(0.062)
Yrs from Last BaseAssess		-0.026	-0.023		-0.023	-0.028
		(0.022)	(0.022)		(0.052)	(0.052)
Ownership	0.130	0.148	0.142	0.203	0.181	0.237
	(0.201)	(0.203)	(0.201)	(0.304)	(0.310)	(0.302)
HCA Ratio	-0.014	-0.021	-0.026	-0.080	-0.149	-0.062
	(0.159)	(0.162)	(0.160)	(0.373)	(0.386)	(0.373)
Pipeline Diameter Large	0.765	0.612	0.835	1.252	1.218	1.433
	(1.048)	(1.053)	(1.047)	(1.759)	(1.707)	(1.748)
Pipeline Diameter Medium	0.035	0.095	0.055	0.239	0.217	0.275
	(0.481)	(0.501)	(0.485)	(0.917)	(0.990)	(0.917)
Average Age	0.001	0.001	0.001	0.001	-0.004	0.002
	(0.005)	(0.005)	(0.005)	(0.009)	(0.010)	(0.009)
Constant	0.862**	0.715*	0.756**	1.442*	1.785**	1.261*
	(0.279)	(0.291)	(0.283)	(0.587)	(0.647)	(0.590)
Pipeline Operator FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Operators	642	642	642	234	234	234
Observations	4,696	4,696	4,696	2,369	2,369	2,369
R^2	0.008	0.007	0.010	0.015	0.012	0.019

Operator-cluster robust standard errors in parentheses; ** p<0.01, * p<0.05, † p<0.1.

Notes. Models 1,2 and 3 includes all observations (full sample), and Models 4, 5 and 6 use the subsample, which excludes pipeline operators that never experienced incidents or that never conducted Integrity Management Program (no Baseline Assessment activities). Per each sample, we run segment regression analysis based on the level of structural complexity: high (Models 2 and 5) and low complexity (Models 3 and 6). Results show that the coefficient of structural complexity is positive and significant except in pipeline operators with high complexity group in subsample (Model 5). We also find that the effect of Baseline Assessment experience is negative and significant under high complexity group (Models 2 and 5).

Table 4.6 Subgroup Analysis Testing Moderation Effects of Complexity (H3)

Sample (Complexity)	Full (All)	Full (High)	Full (Low)	Sub (All)	Sub (High)	Sub (Low)
DV: *Incident CostPerMile*$_{t+1}$	Model	Model	Model	Model	Model	Model
	1	2	3	4	5	6
Complexity	1.034**	0.733	1.323**	2.130**	1.086	2.558**
	(0.332)	(0.478)	(0.456)	(0.584)	(1.051)	(0.700)
BaseAssess Indicator	0.320*	0.304	0.221	0.567*	0.113	0.809*
	(0.155)	(0.229)	(0.217)	(0.265)	(0.370)	(0.396)
BaseAssess Experience	-0.110*	-0.141*	-0.041	-0.170**	-0.169*	-0.162
	(0.047)	(0.058)	(0.080)	(0.062)	(0.076)	(0.102)
Yrs from Last BaseAssess	-0.023	-0.069	0.009	-0.028	-0.202*	0.055
	(0.022)	(0.043)	(0.025)	(0.052)	(0.100)	(0.060)
Ownership	0.142	-0.120	0.499†	0.237	-0.015	0.433
	(0.201)	(0.270)	(0.271)	(0.302)	(0.287)	(0.545)
HCA Ratio	-0.026	0.005	-0.035	-0.062	-0.387	0.185
	(0.160)	(0.332)	(0.171)	(0.373)	(0.482)	(0.549)
Pipeline Diameter Large	0.835	1.084	0.090	1.433	2.880	0.325
	(1.047)	(1.406)	(0.915)	(1.748)	(3.205)	(1.862)
Pipeline Diameter Medium	0.055	0.459	-0.375	0.275	-0.895	0.718
	(0.485)	(0.691)	(0.726)	(0.917)	(1.535)	(1.160)
Average Age	0.001	-0.001	0.003	0.002	-0.019	0.007
	(0.005)	(0.007)	(0.006)	(0.009)	(0.013)	(0.013)
Constant	0.756**	1.268*	0.148	1.261*	3.440**	0.016
	(0.283)	(0.527)	(0.294)	(0.590)	(1.034)	(0.776)
Pipeline Operator FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Operators	642	258	384	234	117	117
Observations	4,696	2,342	2,354	2,369	1,229	1,140
R^2	0.010	0.013	0.019	0.019	0.022	0.036

Operator-cluster robust standard errors in parentheses; ** p<0.01, * p<0.05, † p<0.1.

Notes. Models 1,2 and 3 includes all observations (full sample), and Model 4, 5 and 6 use the subsample, which excludes pipeline operators that never experienced incidents or that never conducted Integrity Management Program (no Baseline Assessment activities). For full sample and subsample respectively, we run segment regression analysis based on the level of structural complexity: high (Models 2 and 5) and low complexity (Models 3 and 6). Results show that the coefficient of structural complexity is positive and significant except in pipeline operators with high complexity group in subsample (Models 2 and 5). We also find that the effect of Baseline Assessment experience is negative and significant under high complexity group (Models 2 and 5).

Robustness Check

We conduct several robustness checks to address concerns regarding the many zeros in the dependent variable, selection bias, and omitted variable bias. We first employ a two-part model to account for mass zeros in the outcome variable, *Incident CostPerMile*. We use this model over other methods (e.g., Heckman selection model, Tobit model) for several reasons First, while the zeros in the Heckman selection model represent censored values, the zeros in the two-part model are real zeros. The two-part model applies to our context, in which we observe non-zero incident cost only if there is an incident. Next, two-part models produce better or more robust estimates than a generalized Tobit model (Belotti et al., 2015). To construct a two-part model, we first build a logit model to predict the likelihood of observing a positive outcome for incident cost. We then construct a generalized linear model (GLM) with a log link conditional on the positive outcome of incident cost per pipeline mileage. Both the first and the second parts contain independent variables, controls, and fixed effects of years, with standard errors clustered in operator-level (Table 4.7). By examining the variations between operators in Model 1, we find that operators with greater complexity ($b_1 = 3.123$, $p < 0.01$, Model 1), no baseline assessment implementation ($b_2 = -0.459$, $p < 0.05$, Model 1), more baseline assessment experience ($b_3 = 0.391$, $p < 0.01$; Model 1), a lower HCA ratio ($b = -0.866$, $p < 0.01$, Model 1), a larger diameter ($b = 2.255$, $p < 0.01$; Model 1), and more aged ($b = 0.021$, $p < 0.01$, Model 1) are likely to incur incidents. Conditioned on the positive outcome of *Incident CostPerMile* in Model 2, we observe that operators with less baseline assessment experience ($b_3 = -0.150$, $p < 0.05$, Model 2, H2 supported) and a greater HCA ratio ($b = 2.441$, $p < 0.01$, Model 2) incur greater incident cost.

Next, to address selection bias with respect to the baseline assessment, we investigate the sensitivity of incident cost to complexity changes between pipeline operators that ever conducted a baseline assessment and those that never did. To investigate sensitivity, we use switching regression, which consists of two stages (Table 4.8). In the first stage, we estimate a probit model (Model 1), which predicts the decision to conduct a baseline assessment. In the second stage, we then estimate ordinary least squares (OLS) regressions on *Incident CostPerMile* between the two groups: operators that conducted a baseline assessment (*Ever BaseAssess* = Yes, Model 2) and those never did (*Ever BaseAssess* = No; Model 3). While both stages include complexity, controls, year fixed effects in common, only the second stage (Models 2 and 3) has a selection correction term (inverse Mills ratio) and operator fixed effects. We find that *Complexity* has a positive effect on incident cost, but only for operators that have conducted a baseline assessment ($b = 1.694$, $p < 0.01$; Model 2), as opposed to operators that have not ($b = -0.733$, $p > 0.10$; Model 3). Two coefficients are significantly different ($Z = 2.877$, $p < 0.01$). Additionally, the insignificant results of the Mills ratio in Models 2 and 3 indicate that omitted factors which may influence the decision to conduct a baseline assessment do not influence future incident cost.

Table 4.7 Robustness Checks: Effects on Likelihood of Incidents and Incident Cost

Sample	Full	Full
DV	*Likelihood of Incident*	*Incident CostPerMile* for Those with Incident
Two-part model	Model	Model
	1	2
Complexity	3.123**	-0.260
	(0.294)	(0.447)
BaseAssess Indicator	-0.459*	-0.077
	(0.228)	(0.306)
BaseAssess Experience	0.391**	-0.150*
	(0.054)	(0.061)
Yrs from Last BaseAssess	-0.027	-0.018
	(0.045)	(0.070)
Ownership	-0.107	0.069
	(0.153)	(0.201)
HCA Ratio	-0.866**	2.441**
	(0.202)	(0.396)
Pipeline Diameter Large	2.255**	0.373
	(0.295)	(0.409)
Pipeline Diameter Medium	1.257**	0.074
	(0.236)	(0.434)
Average Age	0.021**	0.005
	(0.004)	(0.006)
Constant	-2.701**	4.285**
	(0.283)	(0.565)
Pipeline Operator FE	No	No
Year FE	Yes	Yes
Specification	Logit	GLM
# of Operators	642	254
Observations	4,696	1,050
Pseudo R^2	0.304	–
R^2	–	0.141

Operator-cluster standard errors in parentheses; ** p<0.01, * p<0.05, † p<0.1.
Notes. We use a logit model to predict the probability of experiencing incident and employ a generalized linear model (GLM) to estimate the incident consequence (log of incident cost per mile), conditional on the incident.

Table 4.8 Robustness Checks: Sensitivity in Complexity Effects on Incident Cost – (N)ever BaseAssess

Sample	Full	Ever BaseAssess= Yes	Ever BaseAssess = No
DV	Ever BaseAssess (Y/N)	*Incident CostPerMile*$_{t+1}$	*Incident CostPerMile*$_{t+1}$
Switching Regression	Model	Model	Model
	1	2	3
Complexity	4.187**	1.694**	-0.733
	(0.639)	(0.515)	(0.668)
Inverse Mills ratio		-0.134	-0.012
		(0.085)	(0.111)
Ownership	0.878**	0.238	0.367
	(0.314)	(0.196)	(0.339)
HCA Ratio	3.348**	0.365	0.590
	(0.439)	(0.351)	(0.439)
Pipeline Diameter Large	1.559*	1.192	-1.371†
	(0.610)	(1.305)	(0.715)
Pipeline Diameter Medium	1.732**	0.305	-1.916
	(0.457)	(0.541)	(1.512)
Average Age	0.021**	0.000	0.024*
	(0.008)	(0.005)	(0.011)
Constant	1.510*	1.307**	-0.383
	(0.716)	(0.307)	(0.482)
Pipeline Operator FE	No	Yes	Yes
Year FE	Yes	Yes	Yes
Specification	Probit	OLS	OLS
# of Operators	642	464	178
Observations	4,696	4,041	655
Pseudo R^2	0.042	–	–
R^2	–	0.009	0.040

Operator-cluster robust standard errors in parentheses; ** p<0.01, * p<0.05, † p<0.1.

Notes. We examine how sensitivity of *future incident consequence* to *structural complexity* changes between pipeline operators that ever conducted Baseline Assessment (at least once) and those that never did using switching regression. We run a first-stage with a probit model (Model 1), predicting the decision to entry to Baseline Assessment. In a second stage (Models 2 and 3), we then estimate ordinary least squares (OLS) regressions for each group based on the BA entry, respectively. *Ever BA Indicator* represents whether a pipeline operator ever conducted Baseline Assessment. The results show a substantial difference between the effects of structural complexity on *future incident consequence* of pipeline operators that ever-implemented BA versus those that did not: The effect is more statistically significant under pipeline operators with BA. The insignificant results of Mills ratio in Models 2 and 3 indicate that omitted factors that may influence the decision to conduct BA, do not influence *future incident consequence*.

We also take several approaches to mitigate concerns regarding endogeneity from omitted variables. Because we have already examined the issues of mass zeros and baseline assessment, we only report the results of the subsample hereafter, which exclude pipeline operators that never experienced any incidents during the research period and that never conducted baseline assessment activities. This sample consists of 234 pipeline operators with 2,369 operator-year observations (see Appendix C, Table 1). To begin, we additionally include repair experience, the logged amount of pipeline miles repaired by an operator, which may influence the baseline assessment, and incident cost (Table 4.9). We find that the overall results are consistent with previous results, but the coefficient of repair experience is statistically insignificant ($b = -0.017$, $p > 0.10$; Model 3).

Next, we consider enforcement actions carried out by PHMSA. PHMSA initiates an enforcement case if it observes a violation of safety regulations during an inspection. The type of enforcement action is dependent on the seriousness of the violations. While minor problems typically lead to a warning letter, more critical violations may require notices of amendment and monetary penalties (Stafford, 2014). We include four categories of enforcement actions – warning letters, concern letters, notices of amendment, and monetary penalty – in the models (Table A2). While the overall results are consistent with the main results, the coefficients of enforcement actions are insignificant. We note that only warning letters and amendment notices have negative estimates, without statistical significance.

Importantly, we note that the *BaseAssess Indicator* can be endogenous, because some important risk factors influence the baseline assessment decision and the incident cost. To address this potential bias, we consider enforcement regarding IMP (e.g., evaluate whether operators employ appropriate information to assess risks and take needed actions in IMP (Stafford, 2014)) as instrumental variables. Among enforcement actions including warning letters, concern letters, notices of amendment, and monetary penalties issued by PHMSA, we extract enforcement activities exclusive to IMP. We expect that the previous enforcement activities on IMP influence incident cost only through IMP (*BaseAssess*). However, when we conduct two-stage least squares (2SLS) regression with these instrumental variables, the model is underidentified (*Chi-square* = 5.371, $p = 0.717$) and does not empirically meet instrument relevancy. However, we alternatively use generated instruments employing heteroskedastic errors, as suggested by Lewbel (2012). This is a useful approach when external instruments are not available and overidentifying information is needed. The approach considers a subset of regressors, which are not correlated with the product of heteroskedastic errors for identification. This is often the characteristic of statistical models due to unobserved common factors (See Lewbel (2012) for a more detail explanation).

We conduct 2SLS regression with the generated instruments and the standard errors clustered in operator-level. We report the second-stage analyses in Table 4.10. Model 1 concerns operators with all complexity levels, and Models 2 and 3 demonstrate operators with high and low complexity, respectively. Based on the Durbin-Wu-Hausman endogeneity test, we conclude that the *BaseAssess Indicator* is not endogenous and is positively associated with incident cost in Model 2 (p = 0.388) and Model 3 (p = 0.649). While the Hausman test is marginally significant in Model 1 (p = 0.076), the significance and the positive sign of the *BaseAssess Indicator* (b = 1.471, $p <$ 0.05) remain the same, with the main results shown in Table 5. We confirm the validity of instruments through multiple tests of relevance and exclusion restrictions assumptions (Wooldridge, 2002). To test the relevance of instruments, we examine the first stage of 2SLS regression. The F statistics (F = 82.49, Model 1; F = 82.31, Model 2; F = 87.77, Model 3) are greater than the rule of thumb value and within a range indicating strong instruments (Staiger & Stock, 1997). Next, the overidentification test suggests that our instruments satisfy exclusion restriction assumptions (Hansen J (H) = 19.055, p = 0.453, Model 1; H = 19.668, p = 0.415, Model 2; H = 16.824, p = 0.602, Model 3).

One of the interesting controls that we did not include in the main analyses is previous failures, lagged dependent variables, and independent variables (*BaseAssess*). However, including lagged dependent variables leads to dynamic panel structure, and previous literature suggests that fixed effect estimators may be inconsistent when the number of cross-sectional observations is large and the number of time periods is small (Nickell, 1981). To address this concern, we employ a bootstrap-based bias correction procedure for the dynamic panel model (Everaert & Pozzi, 2007). This approach is better than generalized method of moments approaches because it is not vulnerable to the issues of weak instruments and poor small sample property (Bun & Windmeijer, 2010; Everaert & Pozzi, 2007). Our bootstrap procedure contains 250 wild bootstrap iterations with burn-in initialization, which allows general heteroscedasticity. We report the results of including three-year lags of dependent and independent variables in Table 4.11. Model 1 concerns all levels of complexity, and Models 2 and 3 consider high and low complexity groups, respectively. The results show that past *Incident CostPerMile* do not relate to future *Incident CostPerMile*. We find that the main effects of *Complexity* and *BaseAssess* are significant, which is generally consistent with the main results in Table 4.6. However, we note that the effect of *BaseAssess Experience* in low complex groups is also significant ($b = -0.376$, $p < 0.05$, Model 3). We also note that the effects of baseline assessment variables at time t and $t-3$ are statistically significant. Unreported results of one-year lag and two-year lags remain the same.

Table 4.9 Robustness Checks: Repair Activities as Additional Control

Sample (Complexity)	Sub (All)	Sub (High)	Sub (Low)
DV: *Incident CostPerMile*$_{t+1}$	Model	Model	Model
	1	2	3
Complexity	2.148**	1.027	2.643**
	(0.594)	(1.059)	(0.722)
BaseAssess Indicator	0.568*	0.113	0.818*
	(0.265)	(0.371)	(0.398)
BaseAssess Experience	-0.168**	-0.175*	-0.155
	(0.062)	(0.075)	(0.103)
Repair Experience	-0.017	0.047	-0.088
	(0.051)	(0.066)	(0.079)
Years from Last BaseAssess	-0.028	-0.203*	0.053
	(0.052)	(0.100)	(0.060)
Ownership	0.232	-0.014	0.378
	(0.302)	(0.288)	(0.541)
HCA Ratio	-0.065	-0.381	0.168
	(0.374)	(0.479)	(0.547)
Pipeline Diameter Large	1.440	2.888	0.401
	(1.749)	(3.211)	(1.883)
Pipeline Diameter Medium	0.284	-0.903	0.779
	(0.921)	(1.528)	(1.161)
Average Age	0.002	-0.019	0.008
	(0.009)	(0.013)	(0.013)
Constant	0.810	3.413**	0.040
	(0.625)	(1.032)	(0.775)
Pipeline Operator FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
# of Operators	234	117	117
Observations	2,369	1,229	1,140
R^2	0.019	0.023	0.037

Operator-cluster robust standard errors in parentheses; ** p<0.01, * p<0.05, † p<0.1.
Notes. We use a subsample, excluding pipeline operators that never experienced incidents or that never conducted Baseline Assessment. PHMSA measures IMP performance as the number of miles repaired. We additionally control for repair activities, measured as the log of pipelines miles that is identified and repaired.

Table 4.10 Robustness Checks: Endogeneity Test for BaseAssess Indicator

Sample (Complexity)	Sub (All)	Sub (High)	Sub (Low)
DV: *Incident CostPerMile*$_{t+1}$	Model 1	Model 2	Model 3
Complexity	2.075**	0.935	2.521**
	(0.587)	(1.063)	(0.689)
BaseAssess Indicator	1.471*	0.866	1.317†
	(0.585)	(0.573)	(0.763)
BaseAssess Experience	-0.328**	-0.275**	-0.268
	(0.109)	(0.101)	(0.173)
Yrs from Last BaseAssess	0.046	-0.096	0.082
	-0.072	-0.129	-0.066
Ownership	0.252	-0.035	0.445
	(0.304)	(0.290)	(0.538)
HCA Ratio	-0.095	-0.129	0.025
	(0.111)	(0.085)	(0.434)
Pipeline Diameter Large	1.620	2.856	0.427
	(1.736)	(3.145)	(1.814)
Pipeline Diameter Medium	0.272	-0.752	0.654
	(0.889)	(1.524)	(1.117)
Average Age	0.001	-0.019	0.007
	(0.009)	(0.013)	(0.013)
Pipeline Operator FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Underidentification test (*KP LM*)	75.968***	57.334	43.361***
Weak identification test (*CD Wald F*)	16.103	15.305	11.630
Weak identification test (*KP Wald F*)	9.936	19.142	8.937
Hansen J (Overidentification test)	19.055	19.668	16.824
p-value	0.453	0.415	0.602
Hausman test (*p-value*)	0.076	0.388	0.649
# of Operators	234	117	117
Observations	2,369	1,229	1,140

Operator-cluster robust standard errors in parentheses; ** p<0.01, * p<0.05, † p<0.1.
Notes. We use a subsample, excluding pipeline operators that never experienced incidents or that never conducted Baseline Assessment. We use generated instruments with two stage least square regression (2SLS) to test of endogeneity in a baseline assessment indicator suggested by Lewbel (2012). The result supports that baseline assessment indicator is positively associated with baseline assessment indicator.

Table 4.11 Robustness Checks: Lagged Dependent and Independent Variables

Sample (Complexity)	Sub (All)	Sub (High)	Sub (Low)
DV: *Incident CostPerMile*$_{t+1}$	Model 1	Model 2	Model 3
Incident CostPerMile$_{t}$	0.013	0.031	0.010
	(0.034)	(0.041)	(0.043)
Incident CostPerMile$_{t-1}$	-0.013	0.014	-0.043
	(0.034)	(0.048)	(0.041)
Incident CostPerMile$_{t-2}$	0.007	0.016	-0.011
	(0.034)	(0.047)	(0.039)
Complexity	3.599**	1.477	4.118**
	(0.759)	(1.907)	(1.128)
BaseAssess Indicator$_{t}$	1.124**	0.855*	1.416**
	(0.304)	(0.425)	(0.413)
BaseAssess Indicator$_{t-1}$	0.449	0.566	0.272
	(0.279)	(0.402)	(0.442)
BaseAssess Indicator$_{t-2}$	0.525*	0.747†	0.310
	(0.253)	(0.393)	(0.439)
BaseAssess Indicator$_{t-3}$	0.716*	0.259	1.032*
	(0.310)	(0.415)	(0.507)
BaseAsses Experience$_{t}$	-0.280**	-0.236**	-0.376**
	(0.073)	(0.082)	(0.090)
BaseAsses Experience$_{t\ t-1}$	-0.028	-0.044	-0.025
	(0.060)	(0.091)	(0.120)
BaseAsses Experience$_{t\ t-2}$	-0.044	-0.105	0.038
	(0.074)	(0.101)	(0.130)
BaseAsses Experience$_{t\ t-3}$	-0.193**	-0.095	-0.285*
	(0.073)	(0.094)	(0.119)
Yrs from Last BaseAssess	-0.019	-0.077	0.016
	(0.070)	(0.144)	(0.083)
Ownership	0.264	0.015	0.516
	(0.296)	(0.429)	(0.579)
HCA Ratio	-0.027	-0.419	0.368
	(0.168)	(0.603)	(0.747)
Pipeline Diameter Large	-0.485	-0.003	-1.154
	(1.699)	(2.936)	(2.403)
Pipeline Diameter Medium	0.258	0.054	0.559
	(0.945)	(2.600)	(1.729)
Average age	0.005	-0.009	0.007
	(0.010)	(0.017)	(0.020)
Pipeline Operator FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
# of Operators	202	101	101
Observations	1,656	838	818

Bootstrapped standard errors in parentheses; ** p<0.01, * p<0.05, † p<0.1.

Notes. We use bootstrap-corrected fixed-effects estimation and inference in dynamic panel-data models (Everaert & Pozzi, 2007). We set wild bootstrap iterations with burn-in initialization, which allows general heteroscedasticity and does not require an assumption of initial condition. We use a subsample, excluding pipeline operators that never experienced incidents or that never conducted Baseline Assessment.

4.6 Discussion and Conclusion

Recently, organizations in high-hazard industries began to adopt externally mandated, but internally developed, safety management programs (e.g., IMP) to reduce future risk and to enhance safety. However, the academic literature in operations management has focused on more intently on either external aspects (external inspection, penalty) or internal aspects (self-inspection). In this paper, we address organizational learning based upon the safety management program in pipeline operators, where both external and internal agencies are engaged. We show that an increase in structural complexity in a pipeline operator leads to greater costs from subsequent incidents, but organizational learning from the program reduces the costs of future incidents. Importantly, we find that the effect of organizational learning is beneficial when pipeline operators have higher structural complexity in their system. To comprehend our findings, we discuss our results with post-hoc analyses and then provide theoretical contributions and managerial implications.

To better understand the boundary conditions and interpret the findings, we conduct post-hoc analysis by comparing the effects from within-operator variance and between-operator variance. The within effect represents the effect observed over time at the operator level, which indicates the difference based upon an operator's mean-level performance. In contrast, the between effect refers to the effect observed over time across operators in a cross-sectional manner, which specifies the difference in operators' mean-level performance (Ketokivi & McIntosh, 2017; Miller et al., 2018). While the hypotheses and the fixed effects analyses in the previous sections dealt with within-operator variation and within effects, we are also interested in whether major variations in *Complexity*, *BaseAssess* variables, and *Incident CostPerMile* are derived from between-operator variation. To identify these variations, we display the value of intraclass correlation coefficients (ICCs) for each variable in Table 4.3. The ICCs express between-operator variance over total variance, and the sum of within- and between-operator variances. For instance, the ICC of complexity is 0.890, indicating that 89 percent of the variance in complexity is from between operators, while 11 percent is from within operators. We also note that the baseline assessment (BA) and incident cost have a substantial amount of between-operator variance ($ICC_{BA\ indicator}$ = 0.302, $ICC_{BA\ exp}$ = 0.521, $ICC_{Incident\ cost}$ = 0.483).

Since fixed-effects estimators do not denote between effects, we use a hybrid model to incorporate both within and between effects (Allison, 2009; Certo et al., 2017; Schunck, 2013). This approach employs both operator-centered variables and the variables representing operator mean. The former variables capture the within effects, and the latter variables represent the between effects (Certo et al., 2017). Table A3 displays the overall results of the hybrid models. Model 1 concerns operators with all complexity levels, and Models 2 and 3 demonstrate operators with high and low complexity levels, respectively. We confirm that the coefficients of within effects (see Appendix C, Table 3) are consistent with the coefficients of fixed effect models in Table 4.6 (Models 4, 5, and 6). For complexity, our results show that both within effect (b = 2.132, $p < 0.01$) and between effect (b = 1.820, $p < 0.01$) are significant in Model 1. The results indicate that an increase in complexity of a pipeline operator leads to greater incident cost per mile (within effect), and that pipeline operators with higher complexity have greater incident cost per mile than operators with lower complexity (between effect). The results in Models 2 and 3 provide interesting results, indicating that the between effect of complexity is significant when the complexity level is high (b = 6.703, $p < 0.01$, Model 2), and that the within effect of complexity is significant when the complexity level is low (b = 2.504, $p < 0.01$, Model 3). Thus, when the complexity level is high, an increase of complexity in an operator does not influence incident cost per mile, while the relative difference in complexity across operators is positively associated with the incident cost per mile. In contrast, when the complexity level is low, an increase of complexity in an operator does influence the incident cost per mile, while the relative difference in complexity across operators is not associated with the incident cost per mile.

Regarding baseline assessment effects, we have two relevant variables, *BaseAssess Indicator* and *BaseAssess Experience*, in Table A3. For the baseline assessment indicator in Model 1, the results indicate that both within and between effects are statistically significant, but with different directions (b = 0.573, $p < 0.05$, within effect; b = −2.430, $p < 0.01$, between effect). The results show that changing the baseline assessment status of an operator from none to any activities leads to a greater incident consequence (within effect), and that operators with baseline assessment implementation will have less incident consequences than operators without it (between effect). The unexpected positive sign of the within effect is possibly explicated by operators' capability of sensing weak signals (Su et al., 2014). Under the integrity management program (IMP), organizations are more likely to be engaged in operations and to be looking for potential failures. To sense weak signals under IMP, operators will remain attentive to changes in external and internal environments, and they will also take appropriate actions, such as baseline assessment, upon a signal. Therefore, the positive within effect of the baseline assessment indicator does not indicate that the baseline assessment will diminish safety performance, but represents the activities for responsiveness and proactiveness. However, results in Models 2 and 3 suggest that the between effect of the baseline assessment indicator is significant when the complexity level is high (b = −2.591, $p < 0.05$, Model 2), and that the within effect of the baseline assessment indicator is significant when the complexity level is low (b = 0.794, $p < 0.05$, Model 3), suggesting contingency in the effects.

Moving on to the baseline assessment experience, we show that both the within effect and between effect are significant, but with different directions, in Model 1 ($b = -0.170$, $p < 0.01$, between effect; $b = 0.695$, $p < 0.01$, within effect). The results indicate that an increase in experience with baseline assessment, measured as pipeline mileage, leads to a smaller incident consequence (within effect), while pipeline operators with more experience used to have greater incident consequences than operators with less experience (between effect). The results in Models 2 and 3 are similar to the results in Model 1, while suggesting that the within effect of the experience becomes effective when the complexity level is high.

Our research makes two notable theoretical contributions. First, the high-level contribution is that we investigate important arguments about the roles of complexity and organizational learning through an empirical approach. We corroborate the argument that complexity diminishes operational performance, consistent with previous studies (Bode & Wagner, 2015; Bozarth et al., 2009; Wolf, 2001). The results also support the theory that organizational learning can mitigate future incident consequences (Baum & Dahlin, 2007). More importantly, we show that complexity facilitates organizational learning. This is theoretically important because previous studies do not provide a coherent theory and evidence to justify whether complexity fosters or hinders organizational learning (Cohen & Levinthal, 1990; Haunschild & Sullivan, 2002; Leveson et al., 2009; Levitt & March, 1988; Perrow, 1999; Pisano et al., 2001; Rijpma, 1997). Instead, complexity has been considered a double-edge sword in the learning process (Haunschild & Sullivan, 2002; Rijpma, 1997). Our results underpin the argument that complexity heightens the effect of organizational learning.

Second, this study contributes to the growing body of literature on organizational learning in high-hazard industries. Current literature lacks relevant empirical studies about learning in high-hazard organizations, because previous studies have focused on qualitative studies of failures, and the rare events of failures provide only few opportunities to learn (Baum & Dahlin, 2007; Carroll et al., 2002; Labib et al., 2019; Starbuck, 2009). In contrast, we show that high-hazard organizations can learn from an externally mandated, internally developed safety management program (IMP). When we focus on the recent experience from the previous year, we show that learning from the program is effective in reducing subsequent incident consequences. However, our robustness checks examine the evidence of learning from other experiences: inspection with enforcement, repair, and failure. First, we suggest that operators in oil transportation may not learn from recent inspections (see Appendix C, Table 2). While a recent study shows that operators in oil production learn from inspections with enforcement (Mani & Muthulingam, 2019), our results are consistent with previous studies in the pipeline industry that demonstrate federal enforcement actions are poor deterrents in reducing such incidents (Stafford, 2014, 2017). They argue that pipeline operators will comply with regulations only if "the cost of non-compliance exceeds the cost of compliance," implying the Becker's model of crime. This argument also explains our results in terms of why enforcement actions with respect to IMP do not empirically satisfy instrument relevancy. Second, our results suggest that pipeline operators may not learn from experience with repair activities (Table 4.9). Interestingly, a government report points to the possibility that repair activities can increase risk in a pipeline system due to the disturbance in operations (Kowalewski, 2013). Third, we show that operators may not effectively learn from failures by measuring failures from incident cost per mile at one, two, and three years before (Table 4.11). The results coincide with the argument that failure investigations and routine inspections are external events, which will create superficial learning (Haunschild & Rhee, 2004; Marcus, 1988). Similarly, an empirical study of the railroad industry suggests that experiences with past accidents do not reduce future accident costs (Baum & Dahlin, 2007).

Finally, this study offers important managerial implications for regulators, who are responsible for designing policies and regulations. Specifically, we provide insights into whether the integrity management program is effective, where the program is effective, and where we can apply it. First, we support the argument that the IMP has reduced subsequent incident consequences. A previous government report, which deals with data between 1986 and 2012, questions the effectiveness of the program by comparing industry-wide incident consequences and incident frequency before and after the adoption of the program in 2002 (Kowalewski, 2013). In contrast, we show that the program has mitigated future incident consequences by looking at operator-level data with experience with IMP. Specifically, if an operator conducts 10% more baseline assessment over the previous year, given that they conducted the assessment in the previous year as well, the operator will reduce expected incident consequences by 1.6%.[30]

Second, we suggest where the program is effective. Our empirical results show that the program and corresponding activities, including baseline assessment, become effective if a pipeline operator has high structural complexity in its system. Additionally, we further exhibit through post-hoc analysis that the program is fruitful for incidents in high consequence areas (HCAs), but not in non-HCAs. To compare the difference, we examine the influences of base assessment on HCAs (incident cost at HCAs per pipe mileage at HCAs) and non-HCAs (incident cost at non-HCAs per pipe mileage at non-HCAs) in Table 4.12. The results highlight that IMP reduces subsequent incident consequences at HCAs ($b = -0.218$, $p < 0.01$) but not at non-HCAs ($b = -0.067$, $p > 0.10$).

---

[30] Δ*incident cost per pipeline mile* (%) = $e^{-0.170*\ln(1.1)} = 98.4\%$, $100\% - 98.4\% = 1.6\%$ (where *Beta Coefficient* is -0.170 from Model 4 in Table 6).

The results align with the objective of IMP, which is to reduce damages in HCAs by prioritizing resources to HCAs (Kowalewski, 2013). However, we do not observe a spillover effect of IMP on non-HCAs.

Table 4.12 Post-hoc Analysis: Main Effects on Incident Cost Per Mile in HCA and non-HCA

Sample (Complexity)	Sub (All)	Sub (High)	Sub (Low)	Sub (All)	Sub (High)	Sub (Low)
DV	*Incident CostPerMile in HCA$_{t+1}$*			*Incident CostPerMile in non-HCA$_{t+1}$*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Complexity	1.535*	1.281	1.621*	1.818**	1.309	1.982**
	(0.641)	(1.180)	(0.759)	(0.557)	(1.245)	(0.600)
BaseAssess Indicator	0.700*	0.197	1.122**	0.309	-0.097	0.623*
	(0.281)	(0.447)	(0.402)	(0.218)	(0.308)	(0.308)
BaseAssess Experience	-0.218**	-0.190*	-0.264*	-0.067	-0.048	-0.114
	(0.072)	(0.088)	(0.117)	(0.054)	(0.070)	(0.085)
Years from Last BaseAssess	-0.039	-0.188	0.022	0.019	-0.097	0.078
	(0.045)	(0.136)	(0.041)	(0.050)	(0.066)	(0.062)
Ownership	0.280	0.649	-0.150	0.250	-0.401	0.856
	(0.337)	(0.478)	(0.403)	(0.328)	(0.266)	(0.568)
HCA Ratio	-0.026	-0.136	0.338	-0.159	-0.074	-0.477
	(0.093)	(0.092)	(0.372)	(0.129)	(0.081)	(0.385)
Pipeline Diameter Large	-0.682	-1.452	-0.584	1.309	2.684	0.379
	(1.489)	(3.123)	(1.515)	(1.958)	(4.115)	(1.299)
Pipeline Diameter Medium	0.415	0.252	0.362	0.323	-0.278	0.363
	(0.938)	(1.939)	(0.841)	(0.864)	(1.406)	(1.147)
Average Age	-0.004	-0.018	0.006	0.003	-0.005	0.003
	(0.008)	(0.014)	(0.009)	(0.010)	(0.011)	(0.014)
Constant	0.967†	2.473*	0.014	0.389	1.771†	0.313
	(0.537)	(1.214)	(0.513)	(0.610)	(0.961)	(0.749)
Pipeline Operator FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of Operators	234	117	117	234	117	117
Observations	2,369	1,229	1,140	2,369	1,229	1,140
R^2	0.020	0.026	0.044	0.016	0.019	0.037

Operator-cluster robust standard errors in parentheses. ** p<0.01, * p<0.05, † p<0.1.
Notes. We use a subsample, excluding pipeline operators that never experienced incidents or that never conducted Baseline Assessment. We compute *Incident CostPerMile in HCA* as the logged amount of (incident cost in HCA per pipeline mileage in HCA). In a similar vein, we construct *Incident CostPerMile in non-HCA* as the logged amount of (incident cost in non-HCA per pipeline mileage in non-HCA).

Third, our results broadly apply to other contexts in high-hazard industries. IMP is one of the government safety programs referred to as a performance-based regulation (PBR), which has been applied to several other industries. For instance, the Occupational Safety and Health Administration and the Environmental Protection Agency implemented PBRs (e.g., process safety management (PSM), risk management program (RMP)) for chemical plants; the Nuclear Regulatory Commission launched Probabilistic Risk Assessment for nuclear power plants; and the U.S. Coast Guard and the Bureau of Safety and Environmental Enforcement enacted PBRs (e.g., International Safety Management, Safety and Environmental Management System) for shipping companies and offshore oil production facilities (Blanco et al., 2019; Chinander et al., 1998; Coglianese & Olmstead, 2003; DeWolf, 2003; Kleindorfer & Saad, 2005; Kowalewski, 2013; Suzuki, 2014; TRB & NASEM, 2017). However, there are both common and different characteristics between IMP and other PBRs (e.g., PSM, RMP).

The frameworks commonly include organizational procedures of identifying and managing risks, responding to emergencies, and continuously improving safety skills through self-improvement efforts. In contrast, several distinct characteristics of IMP render the context more attractive for investigation. First, granular data related to these efforts are often unavailable in other PBRs (Kowalewski, 2013; Short & Toffel, 2010). Second, IMP articulates the integration of information and requirements more explicitly than other PBRs (DeWolf, 2003; Kowalewski, 2013). For instance, IMP requires documented risk assessment, while PSM and RMP do not require risk assessment as a mandated rule (DeWolf 2003). More specifically, the rule of IMP recommends that the list of risk factors should considered and provides an example of risk ranking methodology. This is a significant difference from other PBRs. Taken together, we propose that other PBRs need to provide guidelines and make their reporting and data more transparent. For instance, a previous study regarding RMP focuses on regulatory compliance outcomes and does not address organizations' internal assessment or practices due to information asymmetry and availability of data (Short & Toffel, 2010).

Limitations and Future Research Directions

This study has limitations, some of which could be mitigated through further research. One limitation is related to the theoretical perspective, and the other concerns the characteristics of data in the analyses. First, during our hypothesis development, we posit that complexity moderates the effect of learning, not that learning moderates the effect of complexity. However, some studies propose that learning nurtures the ability to reduce the cost of complexity (Jacobs & Swink, 2011; Rijpma, 1997). We suggest that future qualitative studies can provide insights in terms of whether complexity moderates the effect of learning or learning moderates the effect of complexity. Next, the annual report data from PHMSA are summary-level, and detailed information is removed. A government report indicates the limitations in the current available data, which does not allow consideration of all of the interactions from various risk factors (Kowalewski, 2013).

We expect our study to spur other researchers to more thoroughly investigate performance-based regulations (PBRs) (e.g., integrity management program) and organizational learning in high-hazard organizations. For instance, one potential study would be to examine PBRs in high-hazard organizations through a qualitative study or a survey method. Specifically, one would address the question of how the adoption of PBRs changes the organizational culture from a compliance culture to a safety culture by incorporating a sense of responsibility, which was highlighted in an existing report (TRB & NASEM, 2017). Relatedly, researchers and practitioners note that it would take a long period of time to make PBR mature by changing the process, system, and culture of pipeline operators (Kowalewski, 2013; Stafford, 2017; Suzuki, 2014).

Next, one would examine the effectiveness of a focused approach on safety management programs. IMP rationalizes that operators should prioritize their resources on high consequence areas rather than applying uniform treatment to all areas. However, investigating when a focused approach is justifiable is an interesting question to study. Finally, we note that among 642 pipeline operators, which consists of 4,696 operator-year observations, 383 operators never had any incidents, and only 259 pipeline operators ever had incidents during our research period, representing 1,050 observations. One approach would be to investigate how operators with no incidents (supposedly high reliability organizations) are different from operators with incidents.

Chapter 5

Conclusion

Managing quality and safety is critical in highly regulated industries because failing to do so can lead to serious negative consequences, such as damage to the environment and loss of human life. One way to improve quality and safety is enhancing organizational focus, emphasis on a specific set of actions or activities. While most of the relevant studies consider that focus is motivated by internal entities, my dissertation extended this perspective by examining focus that are internally and externally driven. The dissertation consisted of three chapters, where each chapter concerns improvements of quality and safety in highly regulated industries: acute-care hospitals, nursing homes, and oil and gas pipeline operators in the United States. I first provide the summary and the contribution of individual chapters and then offer a holistic comparison below.

In chapter two, I studied *internally driven focus* as disproportionate emphasis on a medical specialty in acute-care hospitals. Hospitals are under tremendous pressure to concurrently improve measures of multiple dimensions such as readmission rates and patient satisfaction. Improving these two distinct outcomes by pursuing two improvement approaches - via focus and patient experience - puts conflicting demands on hospital management. These dual goals also pose a considerable challenge for hospital administrators because pursuing focus rests on variation reduction, while improving patient experience increases variation in delivery processes. Using secondary data from 3,027 hospitals, this research demonstrated that focus and patient experience have opposing direct effects on the two measures of performance. Focus has undesirable effects on both the measures while patient experience leads to desirable impacts on both the measures. However, I also demonstrated that the joint effect of focus and patient experience reduces the readmission rates. In contrast, an imbalance between focus and patient experience increases readmission rates and decreases patient satisfaction, both undesirable outcomes from management and patient perspectives. As a set, the study suggested that while managers face challenging tradeoffs in their pursuit to improve multiple dimensions of performance, there is no single magic bullet to improve the two performance measures.

In chapter three, I study *externally driven change in attentional focus* where recurring visits are unannounced while initial visits are announced in advance at nursing homes. Various types of inspections have been used to improve and monitor quality and safety in the manufacturing and service industries. Inspecting agencies make an important choice between two inspection modes: either announce the inspection before arriving at the facility or make an unannounced inspection.

However, despite the debates on the efficacy of these different inspection strategies, there was no coherent theory explains the difference between announced and unannounced inspections, and its immediate and sustained operational performance effects. Using a dataset from accredited nursing homes, the results suggested that both announced and unannounced inspections increase nursing home quality. However, an unannounced inspection leads to more sustained quality performance, while quality performance tends to decline after an announced inspection. Thus, announcing the inspection in advance results in short-term gains but long-term disadvantages. This research provided empirical evidence that announced and unannounced inspections play different roles in affecting immediate and sustained quality performance. I also offered an insight for broad application that unannounced inspections are effective where sustained high-quality performance is critically important, such as the healthcare industry.

In chapter four, I studied *externally driven focus* on a safety management program in oil transportation. High-hazard industries face significant threats of experiencing negative incidents, many of which are highly consequential in their impact (e.g., injuries, fatalities, property loss, and environmental damage). To prevent such incidents, regulatory agencies began to adopt performance-based regulation (Integrity Management Program in oil transportation) to reduce regulator's monitoring burden and cost. Rather than government agencies inspect pipeline operators, the program requires pipeline operators to identify and manage risks, to respond to emergencies, and to continuously improve safety skills by focusing on high consequence areas (HCAs). Using a panel dataset of 642 pipeline operators, this research showed that complexity increases future incident cost but the experience with the program reduces it. Interestingly, complexity heightens the negative relationship between the experience and future incident cost. The program is fruitful for incidents in high consequence areas (HCAs), but not in non-HCAs, which substantiates the intent of the program. This study contributed to the growing literature of organizational learning in high-hazard industries.

To highlight overall findings, I compare the three studies in Table 5.1. These studies commonly addressed the improvements of quality and safety in highly regulated industries through focus. However, they are substantially different in various aspects. First, each study dealt with the different context of focus. While Chapter 2 described *internally driven focus*, Chapters 3 and 4 demonstrated *externally driven focus*. Specifically, Chapter 2 illustrated focus as disproportionate emphasis on a medical specialty. In contrast, Chapter 3 addressed attentional focus on inspection announcement and Chapter 4 studied a mandated safety management program dedicated to high consequence areas. Next, the findings provided implications for different stakeholders: *internally driven focus* for internal entities (hospitals) and *externally driven focus* for external entities (regulators). Chapter 2 suggested hospital administrators to balance focus approach and patient experience practices to purse improvement in multiple dimensions. However, Chapters 3 and 4 offered managerial implications for regulators regarding the use of inspection announcement strategy and the adoption of the safety management program. As a set, my dissertation investigated multiple paths to improve quality and safety in highly regulated industries.

Table 5.1 Dissertation Summary

	Chapter 2	Chapter 3	Chapter 4
Title	The Dark Side of Focus: Is Patient Experience the Cure?	Does Announcing the Visit Matter? An Empirical Examination in US Nursing Homes	Organizational Learning, Complexity, and Safety Management Performance: Evidence from the Oil and Gas Transportation
Industry	Healthcare	Healthcare	Oil and Gas Transportation: Pipeline
Unit of Analysis	Acute-care Hospitals	Accredited Nursing Homes	Pipeline Operators
Data (Years)	Secondary Data (2007-2013)	Secondary Data (2013-2016)	Secondary Data (2004-2017)
Research Question	How do hospitals' execution of a focus strategy and patient experience practices affect multiple dimensions of performance?	Do announced and unannounced inspections lead to an immediate and/or a sustained increase in overall quality performance?	1. Do complexity and organizational learning from a safety management program influence safety performance? 2. Are there contingencies that enhance the relationship between the experience with the program and safety performance?
Focus	Disproportionate emphasis on a medical specialty	Attentional focus on inspections	Focus of the safety management program on high consequence areas
Theoretical Lens	Focus as Emphasis	Attention Based View	Organizational Learning
Research Method	Seemingly Unrelated Regression	Difference-in-Difference; Coarsened Exact Matching;	Fixed Effects Panel Data Model
Implication	Hospitals: Balance between focus strategy and patient experience practices to improve multiple performance simultaneously	Regulators: Use unannounced inspections where sustained high-quality is critically important (e.g., healthcare organizations).	Operators and Regulators: Target safety management program for high consequence areas and high complicated system

Bibliography

Afendulis, C. C., Caudry, D. J., O'Malley, A. J., Kemper, P., Grabowski, D. C., & the THRIVE Research Collaborative. (2016). Green house adoption and nursing home quality. *Health Services Research*, *51*, 454–474.

Aldrich, H. (1979). *Organizations and environments*. Englewood Cliffs, NJ: Prentice-Hall.

Allen, J. E. (2015). *Nursing home federal requirements: Guidelines to surveyors and survey protocols: a user-friendly rendering of the Centers for Medicare and Medicaid's nursing home inspection requirements* (Eighth edition). New York, NY: Springer Publishing Company.

Allison, P. D. (2009). *Fixed effects regression models*. Los Angeles, CA: SAGE.

Al-Ubaydli, O., & McLaughlin, P. A. (2014). RegData: A numerical database on industry-specific regulations for all U.S. industries and federal regulations, *Working Paper*

Anand, G., Gray, J., & Siemsen, E. (2012). Decay, shock, and renewal: Operational routines and process entropy in the pharmaceutical industry. *Organization Science*, *23*(6), 1700–1716.

Anderson, P. (1999). Perspective: Complexity theory and organization science. *Organization Science*, *10*(3), 216–232.

Andritsos, D. A., & Tang, C. S. (2014). Linking process quality and resource usage: An empirical analysis. *Production and Operations Management*, *23*(12), 2163–2177.

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, *90*(430), 431–442.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277–297.

Argote, L., Insko, C. A., Yovetich, N., & Romero, A. A. (1995). Group learning curves: The effects of turnover and task complexity on group performance. *Journal of Applied Social Psychology*, *25*(6), 512–529.

Argote, L., McEvily, B., & Reagans, R. (2003). Managing knowledge in organizations: An integrative framework and review of emerging themes. *Management Science*, *49*(4), 571–582.

Argyris, C. (1977). Double loop learning in organizations. *Harvard Business Review*, *55*(5), 115–125.

Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, *21*(1), 1–42.

Avgerinos, E., & Gokpinar, B. (2017). Team familiarity and productivity in cardiac surgery operations: The effect of dispersion, bottlenecks, and task complexity. *Manufacturing & Service Operations Management*, *19*(1), 19–35.

Axon, R. N., & Williams, M. V. (2011). Hospital readmission as an accountability measure. *JAMA*, *305*(5), 504–505.

Azadegan, A., Patel, P. C., Zangoueinezhad, A., & Linderman, K. (2013). The effect of environmental complexity and environmental dynamism on lean practices. *Journal of Operations Management*, *31*(4), 193–212.

Ball, G., Siemsen, E., & Shah, R. (2017). Do plant inspections predict future quality? The role of investigator experience. *Manufacturing & Service Operations Management*, *19*(4), 534–550.

Banerjee, U. (2017). Shhh... Silent hospital helps healing! *International Journal of Recent Scientific Research*, *8*(10), 21016–21019.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173.

Bar-Yam, Y. (1997). *Dynamics of complex systems*. Reading, MA: Addison-Wesley.

Baum, J. A. C., & Dahlin, K. B. (2007). Aspiration performance and railroads' patterns of learning from train wrecks and crashes. *Organization Science*, *18*(3), 368–385.

Bechel, D. L., Myers, W. A., & Smith, D. G. (2000). Does patient-centered care pay off? *The Joint Commission Journal on Quality Improvement*, *26*(7), 400–409.

Belotti, F., Deb, P., Manning, W. G., & Norton, E. C. (2015). twopm: Two-part models. *The Stata Journal: Promoting Communications on Statistics and Stata*, *15*(1), 3–20.

Bennett, V. M., Pierce, L., Snyder, J. A., & Toffel, M. W. (2013). Customer-driven misconduct: How competition corrupts business practices. *Management Science*, *59*(8), 1725–1742.

Birkinshaw, J., & Gupta, K. (2013). Clarifying the distinctive contribution of ambidexterity to the field of organization studies. *Academy of Management Perspectives*, *27*(4), 287–298.

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *Stata Journal*, *9*(4), 524–546.

Blanco, C. C., Caro, F., & Corbett, C. J. (2019). Managing safety-related disruptions: Evidence from the U.S. nuclear power industry. *Forthcoming in Risk Analysis*.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115–143.

Bode, C., & Wagner, S. M. (2015). Structural drivers of upstream supply chain complexity and the frequency of supply chain disruptions. *Journal of Operations Management*, *36*, 215–228.

Borah, B. J., & Basu, A. (2013). Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. *Health Economics*, *22*(9), 1052–1070.

Borgen, N. T. (2016). Fixed effects in unconditional quantile regression. *Stata Journal*, *16*(2), 403–415.

Boulding, W., Glickman, S. W., Manary, M. P., Schulman, K. A., & Staelin, R. (2011). Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. *The American Journal of Managed Care*, *17*(1), 41–48.

Bowblis, J. R. (2011). Ownership conversion and closure in the nursing home industry. *Health Economics*, *20*(6), 631–644.

Bozarth, C. C., Warsing, D. P., Flynn, B. B., & Flynn, E. J. (2009). The impact of supply chain complexity on manufacturing plant performance. *Journal of Operations Management*, *27*(1), 78–93.

Brooks, J. M., & Ohsfeldt, R. L. (2013). Squeezing the balloon: Propensity scores and unmeasured covariate balance. *Health Services Research*, *48*(4), 1487–1507.

Brush, T., & Karnani, A. (1996). Impact of plant size and focus on productivity: An empirical study. *Management Science*, *42*(7), 1065–1081.

Bun, M. J. G., & Windmeijer, F. (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *The Econometrics Journal*, *13*(1), 95–126.

Cadigan, R. O., Stevenson, D. G., Caudry, D. J., & Grabowski, D. C. (2015). Private investment purchase and nursing home financial health. *Health Services Research*, *50*(1), 180–196.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata*. College Station, TX: Stata Press.

Cao, Q., Gedajlovic, E., & Zhang, H. (2009). Unpacking organizational ambidexterity: Dimensions, contingencies, and synergistic effects. *Organization Science*, *20*(4), 781–796.

Carroll, J. S., Rudolph, J. W., & Hatakenaka, S. (2002). Learning from experience in high-hazard organizations. *Research in Organizational Behavior*, *24*, 87–137.

Castle, N. G., Ferguson, J. C., & Hughes, K. (2009). Humanism in nursing homes: The impact of top management. *Journal of Health & Human Services Administration*, *31*(4), 483–516.

Centers for Medicare and Medicaid Services (CMS). (2012). *Design for nursing home compare five-star quality rating system: Technical users' guide*. Retrieved from https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/.

Centers for Medicare and Medicaid Services (CMS). (2013). *HCAHPS fact sheet*. Retrieved from http://www.hcahpsonline.org/Facts.aspx.

Centers for Medicare and Medicaid Services (CMS). 2014. *Medicare hospital quality chartbook. Performance report on outcome measures.* Retrieved from https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/downloads/medicare-hospital-quality-chartbook-2014.pdf.

Certo, S. T., Withers, M. C., & Semadeni, M. (2017). A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal*, *38*(7), 1536–1556.

Chandrasekaran, A., Senot, C., & Boyer, K. K. (2012). Process management impact on clinical and experiential quality: Managing tensions between safe and patient-centered healthcare. *Manufacturing & Service Operations Management*, *14*(4), 548–566.

Chassin, M. R., Loeb, J. M., Schmaltz, S. P., & Wachter, R. M. (2010). Accountability measures—Using measurement to promote quality improvement. *New England Journal of Medicine*, *363*(7), 683–688.

Chesteen, S., Helgheim, B., Randall, T., & Wardell, D. (2005). Comparing quality of care in non-profit and for-profit nursing homes: A process perspective. *Journal of Operations Management*, *23*(2), 229–242.

Child, J. (1972). Organizational structure, environment and performance: The role of strategic choice. *Sociology*, *6*(1), 1–22.

Chinander, K. R., Kleindorfer, P. R., & Kunreuther, H. C. (1998). Compliance strategies and regulatory effectiveness of performance-based regulation of chemical accident risks. *Risk Analysis*, *18*(2), 135–143.

Choi, T. Y., & Krause, D. R. (2006). The supply base and its complexity: Implications for transaction costs, risks, responsiveness, and innovation. *Journal of Operations Management*, *24*(5), 637–652.

Clark, J. R., & Huckman, R. S. (2012). Broadening focus: Spillovers, complementarities, and specialization in the hospital industry. *Management Science*, *58*(4), 708–722.

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, *100*(5), 1261–1293.

Coglianese, C., & Olmstead, T. (2003). Performance-based regulation: Prospects and limitations in health, safety, and environmental protection. *Administrative Law Review*, *55*, 705–730.

Cohen, M. A., & Kunreuther, H. (2007). Operations risk management: Overview of Paul Kleindorfer's contributions. *Production and Operations Management*, *16*(5), 525–541.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*(1), 128.

Congressional Research Service. (2019). *DOT's Federal Pipeline Safety Program: Background and Key Issues for Congress*. Retrieved from https://fas.org/sgp/crs/misc/R44201.pdf.

Cook, D., Thompson, J. E., Habermann, E. B., Visscher, S. L., Dearani, J. A., Roger, V. L., & Borah, B. J. (2014). From ‘solution shop’ model to ‘focused factory’ in hospital surgery: Increasing care value and predictability. *Health Affairs*, *33*(5), 746–755.

Cowles, C. M. (2007). *Nursing home statistical yearbook, 2007*. McMinnville, OR: Cowles Research Group.

Das, A., Pagell, M., Behm, M., & Veltri, A. (2008). Toward a theory of the linkages between safety and quality. *Journal of Operations Management*, *26*(4), 521–535.

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, *29*(1), 1–19.

Dean, J. W., & Bowen, D. E. (1994). Management theory and total quality: improving research and practice through theory development. *Academy of Management Review*, *19*(3), 392–418.

Dess, G. G., & Beard, D. W. (1984). Dimensions of organizational task environments. *Administrative Science Quarterly*, *29*(1), 52–73.

Detailed hazardous materials incident reports, 49 C.F.R. § 171.16 (2011).

Devaraj, S., Hollingworth, D. G., & Schroeder, R. G. (2004). Generic manufacturing strategies and plant performance. *Journal of Operations Management*, *22*(3), 313–333.

DeWolf, G. B. (2003). Process safety management in the pipeline industry: Parallels and differences between the pipeline integrity management (IMP) rule of the office of Pipeline Safety and the PSM/RMP approach for process facilities. *Journal of Hazardous Materials*, *104*(1), 169–192.

Dhanorkar, S. (2019). Environmental benefits of internet-enabled C2C closed-loop supply chains: A quasi-experimental study of Craigslist. *Management Science*, *65*(2), 660–680.

Dhanorkar, S. S., Siemsen, E., & Linderman, K. W. (2018). Promoting change from the outside: Directing managerial attention in the implementation of environmental improvements. *Management Science*, *64*(6), 2535–2556.

Ding, D. X. (2014). The effect of experience, ownership and focus on productive efficiency: A longitudinal study of U.S. hospitals. *Journal of Operations Management*, *32*(1–2), 1–14.

Ding, X. (David). (2015). The impact of service design and process management on clinical quality: An exploration of synergetic effects. *Journal of Operations Management*, *36*, 103–114.

Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA*, *260*(12), 1743–1748.

Ehlers, L. H., Simonsen, K. B., Jensen, M. B., Rasmussen, G. S., & Olesen, A. V. (2017). Unannounced versus announced hospital surveys: A nationwide cluster-randomized controlled trial. *International Journal for Quality in Health Care*, *29*(3), 1–6.

Epstein, R. M., Fiscella, K., Lesser, C. S., & Stange, K. C. (2010). Why the nation needs a policy push on patient-centered health care. *Health Affairs*, *29*(8), 1489–1495.

Everaert, G., & Pozzi, L. (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control*, *31*(4), 1160–1184.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953–973.

Flynn, B. B., Schroeder, R. G., & Sakakibara, S. (1994). A framework for quality management research and an associated measurement instrument. *Journal of Operations Management*, *11*(4), 339–366.

Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods*, *2*(3), 235–247.

Government Accountability Office (GAO). (2004). *Medicare: CMS needs additional authority to adequately oversee patient safety in hospitals*. Retrieved from https://www.gao.gov/products/GAO-04-850.

Government Accountability Office (GAO). (2006). *Clinical lab quality: CMS and survey organization oversight should be strengthened*. Retrieved from https://www.gao.gov/products/GAO-06-416.

Government Accountability Office (GAO). (2007a). *Drug safety: Preliminary findings suggest weaknesses in FDA's program for inspecting foreign drug manufacturers*. Retrieved from https://www.gao.gov/products/GAO-08-224T.

Government Accountability Office (GAO). (2007b). *Hospital accreditation: Joint Commission on Accreditation of Healthcare Organizations' relationship with its affiliate*. Retrieved from https://www.gao.gov/products/GAO-07-79.

Government Accountability Office (GAO). (2008). *Maritime security: Coast guard inspections identify and correct facility deficiencies, but more analysis needed of program's staffing, practices, and data*. Retrieved from https://www.gao.gov/products/GAO-08-12.

Government Accountability Office (GAO). (2011). *Long-term care hospitals: Differences in their oversight compared to other types of hospitals and nursing homes*. Retrieved from https://www.gao.gov/products/GAO-11-130R.

Government Accountability Office (GAO). (2012a). *Highway infrastructure: Federal-state partnership produces benefits and poses oversight risks*. Retrieved from https://www.gao.gov/products/GAO-12-474.

Government Accountability Office (GAO). (2012b). *Oil and gas management: Interior's reorganization complete, but challenges remain in implementing new requirements*. Retrieved from https://www.gao.gov/products/GAO-12-423.

Giesselmann, M., & Schmidt-Catran, A. (2018). Interactions in Fixed Effects Regression Models. *Social Science Research Network Working Paper*, No. 3227779.

Giordano, L. A., Elliott, M. N., Goldstein, E., Lehrman, W. G., & Spencer, P. A. (2010). Development, implementation, and public reporting of the HCAHPS survey. *Medical Care Research and Review*, *67*(1), 27–37.

Goldstein, S. M., & Iossifova, A. R. (2012). Ten years after: Interference of hospital slack in process performance benefits of quality practices. *Journal of Operations Management*, *30*(1–2), 44–54.

Grabowski, D. C., Afendulis, C. C., Caudry, D. J., O'Malley, A. J., Kemper, P., & the THRIVE Research Collaborative. (2016). The impact of green house adoption on Medicare spending and utilization. *Health Services Research*, *51*, 433–453.

Grabowski, D. C., Angelelli, J. J., & Mor, V. (2004). Medicaid payment and risk-adjusted nursing home quality measures. *Health Affairs*, *23*(5), 243–252.

Gray, J. V., Roth, A. V., & Leiblein, M. J. (2011). Quality risk in offshore manufacturing: Evidence from the pharmaceutical industry. *Journal of Operations Management*, *29*(7–8), 737–752.

Gray, J. V., Siemsen, E., & Vasudeva, G. (2015). Colocation still matters: Conformance quality and the interdependence of R&D and manufacturing in the pharmaceutical industry. *Management Science*, *61*(11), 2760–2781.

Gray, W. B., & Shimshack, J. P. (2011). The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence. *Review of Environmental Economics and Policy*, *5*(1), 3–24.

Greenfield, D., Moldovan, M., Westbrook, M., Jones, D., Low, L., Johnston, B., … Braithwaite, J. (2012). An empirical test of short notice surveys in two accreditation programmes. *International Journal for Quality in Health Care*, *24*(1), 65–71.

Greenfield, D., Travaglia, J., Braithwaite, J., & Pawsey, M. (2007). *Unannounced surveys and tracer methodology: Literature review*. *A Report for the Australian Accreditation Research Network: Examining future health care accreditation research*.

Haunschild, P. R., & Rhee, M. (2004). The role of volition in organizational learning: The case of automotive product recalls. *Management Science*, *50*(11), 1545–1560.

Haunschild, P. R., & Sullivan, B. N. (2002). Learning from complexity: Effects of prior accidents and incidents on airlines' learning. *Administrative Science Quarterly*, *47*(4), 609–643.

He, Z.-L., & Wong, P.-K. (2004). Exploration vs. exploitation: An empirical test of the ambidexterity hypothesis. *Organization Science*, *15*(4), 481–494.

Heidenreich, P. A., Trogdon, J. G., Khavjou, O. A., Butler, J., Dracup, K., Ezekowitz, M. D., … Woo, Y. J. (2011). Forecasting the future of cardiovascular disease in the United States a policy statement from the American Heart Association. *Circulation*, *123*(8), 933–944.

Hendricks, K. B., & Singhal, V. R. (2001). Firm characteristics, total quality management, and financial performance. *Journal of Operations Management*, *19*(3), 269–285.

Hoffman, A. J., & Ocasio, W. (2001). Not all events are attended equally: Toward a middle-range theory of industry attention to external events. *Organization Science*, *12*(4), 414–434.

Huber, G. P. (1991). Organizational Learning: The contributing processes and the literatures. *Organization Science*, *2*(1), 88–115.

Huckman, R. S., & Pisano, G. P. (2006). The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science*, *52*(4), 473–488.

Huckman, R. S., & Zinner, D. E. (2008). Does focus improve operational performance? Lessons from the management of clinical trials. *Strategic Management Journal*, *29*(2), 173–193.

Huesch, M. D. (2009). Learning by doing, scale effects, or neither? Cardiac surgeons after residency. *Health Services Research*, *44*(6), 1960–1982.

Hydari, M. Z., Telang, R., & Marella, W. M. (2019). Saving patient Ryan—Can advanced electronic medical records make patient care safer? *Management Science*. *65*(5), 2041–2059.

Hyer, N. L., Wemmerlöv, U., & Morris Jr., J. A. (2009). Performance analysis of a focused hospital unit: The case of an integrated trauma center. *Journal of Operations Management*, *27*(3), 203–219.

Ibanez, M., & Toffel, M. W. (2019). How scheduling can bias quality assessment: Evidence from food safety inspections. *Forthcoming in Management Science.*

Institute of Medicine (IOM). (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.

Institute of Medicine (IOM). (2009). *Toward health equity and patient-centeredness: Integrating health literacy, disparities reduction, and quality improvement: workshop summary*. Washington, DC: National Academies Press.

Jacobs, M. A., & Swink, M. (2011). Product portfolio architectural complexity and operational performance: Incorporating the roles of learning and fixed assets. *Journal of Operations Management*, *29*(7), 677–691.

Jha, A. K. (2018). To fix the Hospital Readmissions Program, prioritize what matters. *JAMA*, *319*(5), 431–433.

Johnson, M. S. (2018). Regulation by shaming: Deterrence effects of publicizing violations of workplace safety and health laws. *Working Paper*.

Joint Commission on Accreditation of Healthcare Organizations (JCAHO). (2004). *Hospital accreditation standards: Standards, intents*. Oakbrook Terrace, IL: Joint Commission Resources.

Joint Commission on Accreditation of Healthcare Organizations (JCAHO), & Joint Commission Resources (JCR), Inc. (2006). *The Joint Commission's unannounced survey process*. Oakbrook Terrace, IL: Joint Commission Resources.

Joint Commission on Accreditation of Healthcare Organizations (JCAHO), & Joint Commission Resources (JCR), Inc (Eds.). (2007). *Joint Commission unannounced survey case studies*. Oakbrook Terrace, IL: Joint Commission Resources.

Kane, R. L., MacIejewski, M., & Finch, M. (1997). The relationship of patient satisfaction with care and clinical outcomes. *Medical Care*, *35*(7), 714–730.

KC, D. S., & Terwiesch, C. (2011). The effects of focus on performance: Evidence from California hospitals. *Management Science*, *57*(11), 1897–1912.

KC, D., Staats, B. R., & Gino, F. (2013). Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery. *Management Science*, *59*(11), 2435–2449.

Ketokivi, M., & McIntosh, C. N. (2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management*, *52*, 1–14.

Kiefner, J. F., & Trench, C. J. (2001). *Oil pipeline characteristics and risk factors: Illustrations from the decade of construction*. Retrieved from http://www.api.org/~/media/Files/Oil-and-Natural-Gas/PPTS/Other-Files/decadefinal.pdf.

Kim, S.-H. (2015). Time to come clean? Disclosure and inspection policies for green production. *Operations Research*, *63*(1), 1–20.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Forthcoming in Political Analysis.*

Kleindorfer, P. R., & Saad, G. H. (2005). Managing disruption risks in supply chains. *Production and Operations Management*, *14*(1), 53–68.

Klerks, M. C. J. L., Ketelaars, C. A. J., & Robben, P. B. M. (2013). Unannounced, compared with announced inspections: A systematic review and exploratory study in nursing homes. *Health Policy*, *111*(3), 311–319.

Kowalewski, R. (2013). *Pipeline Integrity Management: An evaluation to help improve PHMSA's oversight of performance-based pipeline safety programs*. Retrieved from https://cms.dot.gov/sites/dot.gov/files/docs/IM-PE_Report.pdf.

Labib, A., Hadleigh-Dunn, S., Mahfouz, A., & Gentile, M. (2019). Operationalizing learning from rare events: Framework for middle humanitarian operations managers. *Forthcoming in Production and Operations Management*.

Lee, K. C. L., Sethuraman, K., & Yong, J. (2015). On the hospital volume and outcome relationship: Does specialization matter more than volume? *Health Services Research*, *50*(6), 2019–2036.

Leveson, N., Dulac, N., Marais, K., & Carroll, J. (2009). Moving beyond normal accidents and high reliability organizations: A systems approach to safety in complex systems. *Organization Studies*, *30*(2–3), 227–249.

Levine, D. I., Toffel, M. W., & Johnson, M. S. (2012). Randomized government safety inspections reduce worker injuries with no detectable job loss. *Science*, *336*(6083), 907–911.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(S2), 95–112.

Levitt, B., & March, J. G. (1988). Organizational Learning. *Annual Review of Sociology*, *14*(1), 319–338.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, *30*(1), 67–80.

Lu, S. F., Rui, H., & Seidmann, A. (2018). Does technology substitute for nurses? Staffing decisions in nursing homes. *Management Science*, *64*(4), 1842–1859.

Lu, S. F., & Wedig, G. J. (2012). Clustering, agency costs and operating efficiency: Evidence from nursing home chains. *Management Science*, *59*(3), 677–694.

Madsen, P. M. (2009). These lives will not be lost in vain: Organizational learning from disaster in U.S. coal mining. *Organization Science*, *20*(5), 861–875.

Madsen, P. M., & Desai, V. (2010). Failing to learn? The effects of failure and success on organizational learning in the global orbital launch vehicle industry. *Academy of Management Journal*, *53*(3), 451–476.

Manary, M. P., Boulding, W., Staelin, R., & Glickman, S. W. (2013). The patient experience and health outcomes. *New England Journal of Medicine*, *368*(3), 201–203.

Mani, V., & Muthulingam, S. (2019). Does learning from inspections affect environmental performance? Evidence from unconventional well development in Pennsylvania. *Manufacturing & Service Operations Management*, *21*(1), 177–197.

Marcus, A. A. (1988). Implementing externally induced innovations: A comparison of rule-bound and autonomous approaches. *Academy of Management Journal*, *31*(2), 235–256.

Marcus, A. A., & Nichols, M. L. (1999). On the edge: Heeding the warnings of unusual events. *Organization Science*, *10*(4), 482–499.

Mark, B. A., Harless, D. W., McCue, M., & Xu, Y. (2004). A longitudinal examination of hospital registered nurse staffing and quality of care. *Health Services Research*, *39*(2), 279–300.

Marley, K. A., Collier, D. A., & Meyer Goldstein, S. (2004). The role of clinical and process quality in achieving patient satisfaction in hospitals. *Decision Sciences*, *35*(3), 349–369.

Marsh, L. C., & Cormier, D. R. (2002). *Spline regression models*. Thousand Oaks, CA: Sage Publications.

McDermott, C. M., & Stock, G. N. (2011). Focus as emphasis: Conceptual and performance implications for hospitals. *Journal of Operations Management*, *29*(6), 616–626.

Miller, J. W., Ganster, D. C., & Griffis, S. E. (2018). Leveraging big data to develop supply chain management theory: The case of panel data. *Journal of Business Logistics*, *39*(3), 182–202.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics*, *72*(2), 334–338.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., … Turner, M. B. (2016). Heart disease and stroke statistics—2016 update a report from the American Heart Association. *Circulation*, *133*(4), 447–454.

Nair, A., Nicolae, M., & Narasimhan, R. (2013). Examining the impact of clinical quality and clinical flexibility on cardiology unit performance—Does experiential quality act as a specialized complementary asset? *Journal of Operations Management*, *31*(7–8), 505–522.

National Heart, Lung, and Blood Institute. (2016). What Is Heart Failure? Retrieved from https://www.nhlbi.nih.gov/health/health-topics/topics/hf.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*(6), 1417–1426.

Novak, S., & Eppinger, S. D. (2001). Sourcing by design: Product complexity and the supply chain. *Management Science*, *47*(1), 189–204.

Nyman, J. A. (1993). Testing for excess demand in nursing home care markets. *Medical Care*, *31*(8), 680–693.

Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal*, *18*, 187–206.

Ocasio, W., & Joseph, J. (2017). The attention-based view of great strategies. *Strategy Science*, *3*(1), 289–294.

Peng, D. X., Heim, G. R., & Mallick, D. N. (2014). Collaborative product development: The effect of project complexity on the use of information technology tools and new product development practices. *Production and Operations Management*, *23*(8), 1421–1438.

Perrow, C. (1999). *Normal accidents: Living with high-risk technologies*. Princeton, NJ: Princeton University Press.

Pisano, G. P., Bohmer, R. M. J., & Edmondson, A. C. (2001). Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Science*, *47*(6), 752–768.

Pradhan, R., Weech-Maldonado, R., Harman, J. S., Laberge, A., & Hyer, K. (2013). Private equity ownership and nursing home financial performance: *Health Care Management Review*, *38*(3), 224–233.

Press, M. J., Scanlon, D. P., Ryan, A. M., Zhu, J., Navathe, A. S., Mittler, J. N., & Volpp, K. G. (2013). Limits of readmission rates in measuring hospital quality suggest the need for added metrics. *Health Affairs*, *32*(6), 1083–1091.

Reason, J. T. (1997). *Managing the risks of organizational accidents*. Brookfield, VT: Ashgate.

Reske, K. A., Jenkins, T., Fernandez, C., VanAmber, D., & Hedberg, C. W. (2007). Beneficial effects of implementing an announced restaurant inspection program. *Journal of Environmental Health*, *69*(9), 27–35.

Rijpma, J. A. (1997). Complexity, Tight–Coupling and Reliability: Connecting Normal Accidents Theory and High Reliability Theory. *Journal of Contingencies and Crisis Management*, *5*(1), 15–23.

Ruef, M., & Scott, W. R. (1998). A multidimensional model of organizational legitimacy: Hospital survival in changing institutional environments. *Administrative Science Quarterly*, *43*(4), 877–904.

Schilling, M. A., Vidal, P., Ployhart, R. E., & Marangoni, A. (2003). Learning by doing something else: Variation, relatedness, and the learning curve. *Management Science*, *49*(1), 39–56.

Schmaltz, S. P., Williams, S. C., Chassin, M. R., Loeb, J. M., & Wachter, R. M. (2011). Hospital performance trends on national quality measures and the association with joint commission accreditation. *Journal of Hospital Medicine*, *6*(8), 454–461.

Schunck, R. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *Stata Journal*, *13*(1), 65–76.

Scott, A. (2018). It's Raining, It's Pouring, the Inspector is Snoring: Task Selection in Varying Work Environments. *Social Science Research Network Working Paper*, No. 3212855.

Senot, C., Chandrasekaran, A., & Ward, P. T. (2016a). Role of bottom-up decision processes in improving the quality of health care delivery: A contingency perspective. *Production and Operations Management*, *25*(3), 458–476.

Senot, C., Chandrasekaran, A., Ward, P. T., Tucker, A. L., & Moffatt-Bruce, S. D. (2016b). The impact of combining conformance and experiential quality on hospitals' readmissions and cost performance. *Management Science*, *62*(3), 829–848.

Shaver, J. M. (2019). Interpreting interactions in linear fixed-effect regression models: When fixed-effect estimates are no longer within-effects. *Strategy Science*, *4*(1), 25–40.

Shepherd, D. A., Mcmullen, J. S., & Ocasio, W. (2017). Is that an opportunity? An attention model of top managers' opportunity beliefs for strategic action. *Strategic Management Journal*, *38*(3), 626–644.

Short, J. L., & Toffel, M. W. (2010). Making self-regulation more than merely symbolic: The critical role of the legal environment. *Administrative Science Quarterly*, *55*(3), 361–396.

Short, J. L., Toffel, M. W., & Hugill, A. R. (2016). Monitoring global supply chains. *Strategic Management Journal*, *37*(9), 1878–1897.

Shortell, S. M., Zimmerman, J. E., Rousseau, D. M., Gillies, R. R., Wagner, D. P., Draper, E. A., … Duffy, J. (1994). The performance of intensive care units: Does good management make a difference? *Medical Care*, *32*(5), 508–525.

Siemsen, E., Roth, A., & Oliveira, P. (2010). Common method bias in regression models with linear, quadratic, and interaction effects. *Organizational Research Methods*, *13*(3), 456–476.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*(6), 467–482.

Smith, L., Zheng, N. T., Reilly, K., Kissam, S., Rokoske, F., Barch, D., & Manning, J. (2012). *Nursing home MDS 3.0 quality measures: Final analytic report*. Retrieved from http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/NHQIQualityMeasures.html.

Sosa, E., & Alvarez-Ramirez, J. (2009). Time-correlations in the dynamics of hazardous material pipelines incidents. *Journal of Hazardous Materials*, *165*(1–3), 1204–1209.

Sousa, R., & Voss, C. A. (2002). Quality management re-visited: A reflective review and agenda for future research. *Journal of Operations Management*, *20*(1), 91–109.

Staats, B. R., & Gino, F. (2012). Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Science*, *58*(6), 1141–1159.

Stafford, S. L. (2014). Will additional federal enforcement improve the performance of pipelines in the U.S.? *International Review of Law and Economics*, *37*, 137–146.

Stafford, S. L. (2017). Pipeline Performance and Safety in a Federal System: A Study of Natural Gas Pipeline Enforcement by States in the USA. *Journal of Transport Economics and Policy*, *51*(3), 193–207.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586.

Starbuck, W. H. (2009). Perspective—cognitive reactions to rare events: Perceptions, uncertainty, and learning. *Organization Science*, *20*(5), 925–937.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models, essays in honor of Thomas Rothenberg* (pp. 80–108). New York, NY: Cambridge University Press.

Su, H.-C., & Linderman, K. (2016). An empirical investigation in sustaining high-quality performance. *Decision Sciences*, *47*(5), 787–819.

Su, H.-C., Linderman, K., Schroeder, R. G., & Van de Ven, A. H. (2014). A comparative case study of sustaining quality as a competitive advantage. *Journal of Operations Management*, *32*(7–8), 429–445.

Suzuki, A. (2014). Managing the Fukushima Challenge: Managing the Fukushima Challenge. *Risk Analysis*, *34*(7), 1240–1256.

Transportation Research Board (TRB), & the National Academy of Sciences, Engineering, and Medicine (NASEM). (2017). *Designing Safety Regulations for High-Hazard Industries*. Washington, DC: National Academy of Sciences.

Transportation of hazardous liquids by pipeline, 49 C.F.R. § 195 (2006).

Transportation of hazardous liquids by pipeline, 49 C.F.R. pt. 195 (2010).

Trench, C. J. (2003). *The US oil pipeline industry's safety performance*. Retrieved from http://puc.sd.gov/commission/dockets/hydrocarbonpipeline/2007/Hp07-001/121107performance.pdf.

Tsikriktsis, N. (2007). The effect of operational performance and focus on profitability: A longitudinal study of the U.S. airline industry. *Manufacturing & Service Operations Management*, *9*(4), 506–517.

Venkatraman, N. (1989). The concept of fit in strategy research: Toward verbal and statistical correspondence. *Academy of Management Review*, *14*(3), 423–444.

Vickery, S. K., Koufteros, X., Dröge, C., & Calantone, R. (2016). Product modularity, process modularity, and new product introduction performance: Does complexity matter? *Production and Operations Management*, *25*(4), 751–770.

Wagner, L. M., McDonald, S. M., & Castle, N. G. (2012a). Impact of voluntary accreditation on deficiency citations in U.sS nursing homes. *The Gerontologist*, *52*(4), 561–570.

Wagner, L. M., McDonald, S. M., & Castle, N. G. (2012b). Relationship between nursing home safety culture and Joint Commission Accreditation. *Joint Commission Journal on Quality and Patient Safety*, *38*(5), 207–215.

Wernerfelt, B., & Montgomery, C. A. (1988). Tobin's q and the importance of focus in firm performance. *The American Economic Review*, *78*(1), 246–250.

Westbrook, K. W., Babakus, E., & Grant, C. C. (2014). Measuring patient-perceived hospital service quality: Validity and managerial usefulness of HCAHPS scales. *Health Marketing Quarterly*, *31*(2), 97–114.

Westphal, J. D., Gulati, R., & Shortell, S. M. (1997). Customization or conformity? An institutional and network perspective on the content and consequences of TQM adoption. *Administrative Science Quarterly*, *42*(2), 366.

Wiengarten, F., Fan, D., Lo, C. K. Y., & Pagell, M. (2017). The differing impacts of operational and financial slack on occupational safety in varying market conditions. *Journal of Operations Management*, *52*, 30–45.

Wolf, F. G. (2001). Operationalizing and testing normal accident theory in petrochemical plants and refineries. *Production and Operations Management*, *10*(3), 292–305.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*(298), 348–368.

Zhang, G. P., & Xia, Y. (2013). Does quality still pay? A reexamination of the relationship between effective quality management and firm performance. *Production and Operations Management*, *22*(1), 120–136.

Appendix

Appendix A. The Dark Side of Focus (Chapter 2)

Table 1 Patient Experience Measure (HCAHPS)

Category	Description	Comments*	Answer**
COM1	General communications with nurses	Communication with	A/U/SN
	Q1. Nurses treated you with courtesy and respect		
	Q2. Nurses listened carefully to you		
	Q3. Nurses explained things clearly		
COM2	General communications with doctors	Communication with	A/U/SN
	Q1. Doctors treated you with courtesy and respect		
	Q2. Doctors listened carefully to you		
	Q3. Doctors explained things clearly		
COM3	Explanation of medicines	Communication with	A/U/SN
	Q1. Staffs told you what new medicine was for		
	Q2. Staffs described possible side effect of new medicine		
COM4	Discharge Information	Communication with	Y/N
	Q1. Staffs told you would have the help you needed		
	Q2. You got written instructions about what symptoms or		
RES1	Staffs' responsiveness	Responsiveness to patient	A/U/SN
	Q1. You got help as soon as wanted after pressing the		
	Q2. You got help as soon as wanted to use the restroom		
RES2	Pain management	Responsiveness to patient	A/U/SN
	Q1. Your pain was well controlled		
	Q2. Hospital staffs did everything to help you with your		
ENV1	Q. Room and bathroom kept clean	Hospital environment	A/U/SN
ENV2	Q. Area around room quiet at night	Hospital environment	A/U/SN

*We follow the categorization of Westbrook et al. (2014): communication, responsiveness, and hospital environment.
**A/U/SN: Always, Usually, Sometimes or Never; Y/N: Yes, No

Table 2 Clinical Process Quality Measure

	Label	Measure Short Name (adopted from TJC)
HF	HF-2	Evaluation of LVS Function
	HF-3	ACEI or ARB for LVSD
AMI	AMI-2	Aspirin Prescribed at Discharge
	AMI-8a	Primary PCI Received Within 90 Minutes of Hospital Arrival
PN	PN-3b	Blood Culture Before First Antibiotic
	PN-6	Initial Antibiotic Selection for CAP in Immunocompetent Patient

Notes. We exclude retired measure items during our data period such as 'Adult Smoking Cessation Advice' (HF-4/AMI-4/PN-4). We also drop an item about discharge instruction in heart failure care (HF-1) because it is not capturing process of care but outcome of care (accountability issue) (Chassin et al., 2010).

Table 3 Two-Stage Least Squares (2SLS) Regression Results: Readmission Rates

	Model
	2SLS
Focus	0.218**
	(0.0912)
PatExp	-0.0990***
	(0.0200)
Time	-0.842***
	(0.0380)
For-profit	0.201***
	(0.0295)
Government	-0.0238
	(0.0361)
Teaching	0.0756***
	(0.0271)
Location	-0.0286
	(0.0300)
Hospital size	0.144***
	(0.0211)
Bed occ. rate	0.141***
	(0.0152)
CMI	-0.263***
	(0.0408)
Nursing intensity	0.0198
	(0.0166)
CPQ (*HF*)	-0.0201*
	(0.0121)
Constant	0.376***
	(0.0403)
Time FE	Yes
State FE	Yes
Observations	5,862
R^2	0.353

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Notes. Based on the Durbin-Wu-Hausman endogeneity test (χ^2 = 0.17; p = 0.678), we conclude that focus is not endogenous in the readmission rates model.

Table 4 Subgroup Analysis of Resource Utilization on Readmission Rates

	High Bed Occupancy			Low Bed Occupancy		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Focus	0.228***		0.216***	0.091***		0.088***
	(0.029)		(0.031)	(0.019)		(0.018)
PatExp		-0.118***	-0.083*		-0.071†	-0.067†
		(0.033)	(0.033)		(0.040)	(0.037)
For-profit	0.259***	0.251***	0.226***	0.192***	0.165**	0.164**
	(0.041)	(0.045)	(0.044)	(0.052)	(0.054)	(0.055)
Government	0.127*	0.193***	0.115*	-0.085†	-0.070	-0.082†
	(0.052)	(0.054)	(0.050)	(0.046)	(0.043)	(0.045)
Teaching	0.026	0.063	0.022	0.020	0.017	0.015
	(0.039)	(0.041)	(0.041)	(0.048)	(0.047)	(0.047)
Location	0.023	0.007	-0.004	0.028	0.018	0.017
	(0.084)	(0.087)	(0.082)	(0.031)	(0.031)	(0.030)
Hospital size	0.131***	0.065†	0.107**	0.131***	0.092**	0.104**
	(0.038)	(0.035)	(0.038)	(0.028)	(0.035)	(0.034)
CMI	-0.150***	-0.218***	-0.137***	-0.238***	-0.260***	-0.229***
	(0.027)	(0.032)	(0.028)	(0.031)	(0.036)	(0.034)
Nursing intensity	0.018	0.011	0.019	-0.036*	-0.044*	-0.035†
	(0.019)	(0.021)	(0.019)	(0.018)	(0.018)	(0.018)
CPQ (*HF*)	-0.006	-0.001	-0.001	-0.032†	-0.030†	-0.027
	(0.015)	(0.015)	(0.014)	(0.017)	(0.017)	(0.017)
Constant	-0.605***	-0.730***	-0.577***	0.810***	0.733***	0.805***
	(0.091)	(0.097)	(0.091)	(0.039)	(0.033)	(0.039)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,931	2,931	2,931	2,931	2,931	2,931
F	36.99	33.65	36.9	33.13	32.84	32.90
R^2	0.432	0.409	0.4355	0.405	0.403	0.408

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Notes. The effect of focus on readmission rates is significantly greater in the high occupancy group than low occupancy group (p < 0.001).

Table 5 Subgroup Analysis of Teaching Status on Readmission Rates

	Teaching			Non-Teaching		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Focus	0.222***		0.208***	0.123***		0.120***
	(0.041)		(0.039)	(0.019)		(0.020)
PatExp		-0.114**	-0.065*		-0.082*	-0.076*
		(0.036)	(0.030)		(0.036)	(0.035)
For-profit	0.207***	0.188***	0.182**	0.257***	0.227***	0.224***
	(0.062)	(0.057)	(0.059)	(0.039)	(0.042)	(0.044)
Government	0.120†	0.197**	0.110	-0.035	-0.017	-0.033
	(0.073)	(0.073)	(0.071)	(0.036)	(0.033)	(0.035)
Teaching	-0.061	-0.043	-0.079	0.052	0.037	0.037
	(0.069)	(0.071)	(0.065)	(0.045)	(0.044)	(0.044)
Location	0.124**	0.082*	0.107*	0.116***	0.068*	0.087**
	(0.045)	(0.041)	(0.043)	(0.026)	(0.033)	(0.033)
Hospital size	0.207***	0.188***	0.182**	0.257***	0.227***	0.224***
	(0.062)	(0.057)	(0.059)	(0.039)	(0.042)	(0.044)
CMI	-0.161***	-0.239***	-0.156***	-0.257***	-0.282***	-0.243***
	(0.044)	(0.046)	(0.043)	(0.030)	(0.037)	(0.034)
Nursing	0.020	0.000	0.018	-0.023	-0.031†	-0.022
intensity	(0.029)	(0.030)	(0.029)	(0.016)	(0.016)	(0.016)
CPQ (*HF*)	0.000	0.004	0.003	-0.023	-0.023	-0.018
	(0.018)	(0.019)	(0.018)	(0.017)	(0.018)	(0.018)
Constant	-0.683***	-0.793***	-0.668***	0.618***	0.512***	0.619***
	(0.082)	(0.085)	(0.082)	(0.045)	(0.035)	(0.045)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,939	1,939	1,939	3,923	3,923	3,923
F	25.96	23.95	25.72	44.00	43.01	43.79
R^2	0.449	0.429	0.4511	0.402	0.396	0.405

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Notes. The effect magnitude of focus on readmission rates is significantly greater in teaching group than non-teaching group (p < 0.05).

Table 6 Subgroup Analysis of Resource Utilization on Patient Satisfaction

	High Bed Occupancy			Low Bed Occupancy		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Focus	-0.224***		-0.098**	-0.128***		-0.093***
	(0.058)		(0.035)	(0.034)		(0.023)
PatExp		0.912***	0.897***		0.891***	0.887***
		(0.038)	(0.037)		(0.070)	(0.068)
For-profit	-0.501***	-0.157**	-0.144**	-0.556***	-0.172***	-0.171***
	(0.109)	(0.051)	(0.052)	(0.068)	(0.037)	(0.037)
Government	-0.160**	-0.083*	-0.049	-0.046	-0.094**	-0.081**
	(0.056)	(0.041)	(0.038)	(0.065)	(0.029)	(0.030)
Teaching	-0.031	-0.011	0.007	-0.044	0.017	0.019
	(0.067)	(0.036)	(0.036)	(0.046)	(0.026)	(0.025)
Location	-0.061	0.224***	0.230***	0.013	0.163***	0.165***
	(0.073)	(0.047)	(0.046)	(0.053)	(0.028)	(0.028)
Hospital size	-0.183***	0.085***	0.066***	-0.361***	0.009	-0.004
	(0.032)	(0.018)	(0.017)	(0.034)	(0.030)	(0.031)
CMI	0.243***	0.145***	0.108***	0.262***	0.192***	0.160***
	(0.042)	(0.019)	(0.021)	(0.023)	(0.021)	(0.016)
Nursing	0.044*	0.046**	0.043**	0.052†	0.050*	0.041*
intensity	(0.019)	(0.015)	(0.015)	(0.029)	(0.020)	(0.019)
CPQ (*All*)	0.137***	0.053***	0.049***	0.171***	0.079***	0.075***
	(0.030)	(0.013)	(0.014)	(0.029)	(0.019)	(0.018)
Constant	0.143	-0.121†	-0.193**	-0.210***	-0.101***	-0.179***
	(0.091)	(0.069)	(0.066)	(0.045)	(0.026)	(0.030)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,931	2,931	2,931	2,931	2,931	2,931
F	34.84	233.69	238.86	23.46	152.82	154.55
R^2	0.417	0.828	0.833	0.325	0.758	0.764

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Notes. The effect magnitude of focus on patient satisfaction is significantly greater in the high occupancy group than low occupancy group (p = 0.07).

Table 7 Subgroup Analysis of Teaching Status on Patient Satisfaction

	Teaching			Non-Teaching		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Focus	-0.309***		-0.118**	-0.114***		-0.080***
	(0.064)		(0.038)	(0.031)		(0.021)
PatExp		0.888***	0.860***		0.958***	0.955***
		(0.053)	(0.047)		(0.062)	(0.061)
For-profit	-0.481***	-0.147*	-0.142†	-0.574***	-0.146***	-0.143***
	(0.100)	(0.073)	(0.075)	(0.079)	(0.034)	(0.036)
Government	-0.123*	-0.046	0.003	-0.064	-0.108***	-0.098***
	(0.061)	(0.039)	(0.040)	(0.049)	(0.026)	(0.026)
Teaching	-0.048	0.180***	0.202***	-0.017	0.163***	0.162***
	(0.096)	(0.052)	(0.052)	(0.056)	(0.027)	(0.027)
Location	-0.134**	0.090**	0.076**	-0.343***	0.027	0.014
	(0.042)	(0.029)	(0.028)	(0.028)	(0.022)	(0.023)
Hospital size	-0.006	0.099***	0.091***	0.004	0.132***	0.124***
	(0.042)	(0.021)	(0.023)	(0.024)	(0.013)	(0.013)
CMI	0.176***	0.154***	0.108***	0.298***	0.158***	0.133***
	(0.030)	(0.026)	(0.025)	(0.026)	(0.016)	(0.015)
Nursing	0.003	0.038*	0.028	0.044†	0.050**	0.045**
intensity	(0.026)	(0.016)	(0.017)	(0.023)	(0.018)	(0.017)
CPQ (*All*)	0.123***	0.051†	0.048	0.157***	0.062***	0.058***
	(0.027)	(0.029)	(0.029)	(0.024)	(0.013)	(0.012)
Constant	0.214**	0.027	-0.047	-0.179***	-0.134***	-0.205***
	(0.079)	(0.062)	(0.070)	(0.053)	(0.024)	(0.025)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,939	1,939	1,939	3,923	3,923	3,923
F	26.38	128.80	132.4	33.14	253.03	254.39
R^2	0.453	0.802	0.8088	0.336	0.794	0.798

Clustered Standard Errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05, † p<0.1.

Notes. The effect magnitude of focus on patient satisfaction is significantly greater in teaching group than non-teaching group (p < 0.01).

Appendix B. Does Announcing the Visit Matter (Chapter 3)

Table 1 Quality Measures and Description

Measure	Description (% of)	All	Long	Short
long 401	Long stay residents whose need for help with ADLs has increased‡	X	X	
long 402	Long stay residents who self-report moderate to severe pain†	X	X	
long 403	High-risk long stay residents with pressure ulcers‡	X	X	
long 406	Long stay residents with a catheter inserted and left in their bladder†	X	X	
long 407	Long stay residents with a urinary tract infection‡	X	X	
long 409	Long stay residents who were physically restrained‡	X	X	
long 410	Long stay residents experiencing one or more falls with major injury‡	X	X	
long 419	Long stay residents who received an antipsychotic medication‡	X	X	
short 424	Short stay residents who self-report moderate to severe pain‡	X		X
short 425	Short stay residents with pressure ulcers that are new or worsened†	X		X
short 434	Short stay residents who newly received an antipsychotic medication‡	X		X

†Model-based risk-adjusted (using resident-level covariates that increase the risks of an outcome)
‡Sampling-based risk-adjusted (excluding residents whose outcomes are unavoidable or not under nursing homes' control)

Table 2 Sample Reduction and Matching

	Original Sample			Reduced Sample for DV Availability			Reduced Sample for One Inspection			Reduced Sample After Matching	
Variables	All	TJC	Non-TJC	All	TJC	Non-TJC	All	TJC	Non-TJC	TJC	Non-TJC
For-profit	73.8%	80.8%	73.4%	74.8%	79.2%	74.3%	74.9%	82.8%	74.3%	83.8%	83.8%
Not-for-profit	22.4%	18.0%	22.6%	22.1%	19.9%	22.3%	21.9%	16.6%	22.3%	15.9%	15.9%
Chain	58.5%	61.7%	58.3%	60.2%	62.4%	60.0%	60.1%	62.0%	60.0%	62.4%	62.4%
Location	71.1%	90.6%	69.9%	83.0%	92.7%	82.0%	82.8%	92.3%	82.0%	93.0%	93.0%
TJC State	33.8%	40.7%	33.4%	33.3%	38.9%	32.7%	33.3%	41.0%	32.7%	40.4%	40.4%
Size (bed)	111.2	144.1	109.3	147.3	159.1	146.1	147.1	159.5	146.1	160.5	155.1
	(59.3)	(62.0)	(58.5)	(63.7)	(64.1)	(63.6)	(63.8)	(65.7)	(63.6)	(65.9)	(59.2)
Medicare (%)	15.3	20.4	15.0	16.1	18.5	15.9	16.1	19.0	15.9	19.0	17.0
	(13.0)	(15.7)	(12.8)	(9.0)	(10.9)	(8.7)	(9.0)	(11.5)	(8.7)	(11.60)	(8.63)
Medicaid (%)	60.4	56.3	60.7	61.3	59.0	61.5	61.3	58.5	61.5	58.5	60.5
	(21.5)	(20.7)	(21.6)	(15.1)	(14.7)	(15.1)	(15.1)	(15.6)	(15.1)	(15.7)	(14.6)
Acuity Index	12.1	12.2	12.1	12.4	12.3	12.4	12.4	12.4	12.4	12.4	12.4
	(1.3)	(1.3)	(1.3)	(1.0)	(0.8)	(1.0)	(1.0)	(0.8)	(1.0)	(0.8)	(0.9)
OP Margin (%)	-7.5	-1.7	-14.2	-2.3	-1.2	-2.4	-2.3	-1.1	-2.4	-1.0	-0.5
	(67.6)	(12.8)	(31.2)	(27.4)	(11.5)	(28.5)	(27.7)	(12.2)	(28.5)	(12.3)	(17.6)
Observations	13,324	733	12,591	4,926	452	4,474	4,812	337	4,474	331	331

Standard deviations in the parentheses.

Table 3 Bootstrap-based Bias Correction

	(1) Announced Inspection			(2) Unannounced Inspection		
	All residents Quality	Long-stay Quality	Short-stay Quality	All residents Quality	Long-stay Quality	Short-stay Quality
	1	2	3	1	2	3
2 *Quarters Prior*	1.613	2.260	-0.155	9.690	6.291	3.491
	(5.966)	(3.934)	(2.552)	(7.093)	(6.447)	(2.723)
1 *Quarter Prior*	1.243	-1.003	2.221	7.812	7.258	0.939
	(4.441)	(3.703)	(2.579)	(7.801)	(6.531)	(3.813)
Inspection Visit	11.293*	8.210*	2.716	9.928†	0.992	9.453**
	(4.883)	(4.085)	(2.112)	(6.000)	(5.060)	(2.745)
1 *Quarter After*	-3.396	-4.694	2.198	7.302	2.205	5.666†
	(4.845)	(4.157)	(2.076)	(6.492)	(5.857)	(3.132)
2 *Quarters After*	-0.930	1.233	-1.938	9.665	9.066	1.159
	(4.871)	(4.035)	(2.375)	(6.026)	(5.657)	(2.661)
DV (*t*-1)	0.633**	0.582**	0.605**	0.643**	0.583**	0.612**
	(0.014)	(0.015)	(0.011)	(0.014)	(0.016)	(0.012)
Nursing FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
# of NHs	552	552	552	441	441	441
Observations	8,280	8,280	8,280	6,615	6,615	6,615

Standard errors in parentheses; ** p<0.01, * p<0.05, †p<0.1; Used wild bootstrap.

Table 4 TJC and State Agency (SA) Inspections Visits

	(1) Announced Inspection by TJC			(2) Unannounced Inspection by TJC		
	All residents Quality	Long-stay Quality	Short-stay Quality	All residents Quality	Long-stay Quality	Short-stay Quality
	1	2	3	1	2	3
2 *Quarters Prior*	1.759	2.020	-0.209	8.950	5.504	3.656
	(5.024)	(3.913)	(2.942)	(6.676)	(6.046)	(3.134)
1 *Quarter Prior*	0.994	-1.199	2.098	8.108	7.418	1.102
	(4.711)	(4.012)	(2.500)	(7.815)	(6.588)	(3.720)
TJC's Inspection Visit	11.285*	7.923*	3.189	10.759†	1.616	9.690**
	(4.747)	(3.815)	(2.316)	(6.138)	(5.185)	(3.017)
1 *Quarter After*	-2.088	-4.028	2.283	8.999	2.796	6.761*
	(5.060)	(3.996)	(2.560)	(6.786)	(6.036)	(2.862)
2 *Quarters After*	0.095	1.528	-1.325	10.781†	9.096†	2.217
	(4.801)	(3.919)	(2.393)	(6.255)	(5.200)	(3.072)
SA's Unannounced Visit	2.304	1.360	1.012	-0.466	0.055	-0.443
	(1.844)	(1.535)	(0.908)	(2.144)	(1.770)	(1.014)
DV (*t*-1)	0.516**	0.470**	0.489**	0.525**	0.472**	0.496**
	(0.012)	(0.013)	(0.010)	(0.013)	(0.014)	(0.012)
Constant	349.951**	275.132**	103.400**	344.986**	274.993**	102.950**
	(8.694)	(6.609)	(2.529)	(9.649)	(7.516)	(2.879)
Nursing FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Nursing Homes	552	552	552	441	441	441
Observations	8,280	8,280	8,280	6,615	6,615	6,615
R^2	0.282	0.242	0.242	0.299	0.249	0.250

Clustered Standard Errors in parentheses; ** p<0.01, * p<0.05, †p<0.1.

Table 5 Conditions of announced and unannounced inspections

Inspection	Prior notice	Accreditation Program MCLTC	NCC (LTC)
Announced Visit	30 days	Initial inspection	Initial inspection
Announced Visit	7 days	(1) Recurring inspection for freestanding nursing homes (2) First inspection after converting from MCLTC to NCC program	
Unannounced Visit	0 days	Recurring inspection for hospital affiliated nursing homes	Recurring inspection

Notes. To fully comprehend the conditions of announced and unannounced inspections, we delineate two types of TJC accreditation programs for nursing homes: nursing care center (NCC) accreditation program, previously known as long-term care (LTC) accreditation program, and Medicare and Medicaid certification-based long-term care (MCLTC) accreditation program. Both programs provide the framework of TJC standards for nursing homes, including leader's role in promoting safety and quality, risk and safety management, education for staff and residents. While the NCC program evaluates all TJC standards for care quality and safety, the MCLTC program contains only a subset of TJC standards, which are not evaluated during the state survey agency's inspection. Relatedly, the MCLTC accreditation program costs less than the NCC program and typically involves one-day inspection. TJC introduced MCLTC program in 2003 for those nursing homes with limited financial resources but it expired in 2013–2014. While TJC have conduced inspection on the unannounced basis since 2006, some inspections were announced. This table gives an overview of the conditions for announced and unannounced inspections. First, if a nursing home undergoes its first TJC inspection, the schedule of inspection is announced at least thirty days before the actual visit. Second, when a nursing home is under the MCLTC program and it is not part of a hospital, there will be seven-day notice before the inspection visit. Third, if a nursing home shifts its accreditation from the MCLTC program to the NCC program due to the MCLTC program retirement, the first subsequent inspection is announced seven days prior to the visit. However, all other inspections are unannounced. We note that, recently, there were no 7-day announced inspections in TJC's nursing home accreditation program, since the MCLTC program was retired from TJC.
(https://www.jointcommission.org/assets/1/18/Unannounced_Survey_Process_9_12.pdf).

Table 6 Effects of 30-day and 7-day Announced Inspection Visits

	All residents Quality 1	Long-stay Quality 2	Short-stay Quality 3
2 *Quarters Prior*	5.196	-1.075	6.179
	(8.138)	(7.479)	(4.666)
1 *Quarter Prior*	1.692	1.477	0.347
	(8.204)	(7.005)	(4.187)
Announced Visit	7.259	6.084	1.360
(*30-day notice*)	(8.359)	(6.283)	(4.292)
1 *Quarter After*	-3.830	-10.874†	7.557†
	(8.177)	(6.529)	(4.183)
2 *Quarters After*	1.932	3.509	-1.674
	(8.090)	(6.780)	(3.808)
2 *Quarters Prior*	0.021	3.720	-3.540
	(6.327)	(4.480)	(3.689)
1 *Quarter Prior*	0.553	-2.603	2.937
	(5.760)	(4.915)	(3.058)
Announced Visit	13.297*	8.838†	4.085
(*7-day notice*)	(5.757)	(4.785)	(2.705)
1 *Quarter After*	-1.255	-0.524	-0.469
	(6.448)	(5.025)	(3.202)
2 *Quarters After*	-0.731	0.720	-1.254
	(5.992)	(4.833)	(3.015)
DV (*t*-1)	0.516**	0.471**	0.490**
	(0.012)	(0.012)	(0.010)
Constant	349.879**	275.115**	103.295**
	(8.689)	(6.603)	(2.528)
Nursing FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
# of Nursing Homes	552	552	552
Observations	8,280	8,280	8,280
R^2	0.282	0.242	0.242

Cluster-robust standard errors are given in parentheses; ** p<0.01, * p<0.05, †p<0.1.
Notes. 30-day notice (n = 76), 7-day notice (n = 145)

Table 7 Immediate Effects of (Un)Announced in Quantile Regression: All resident Quality

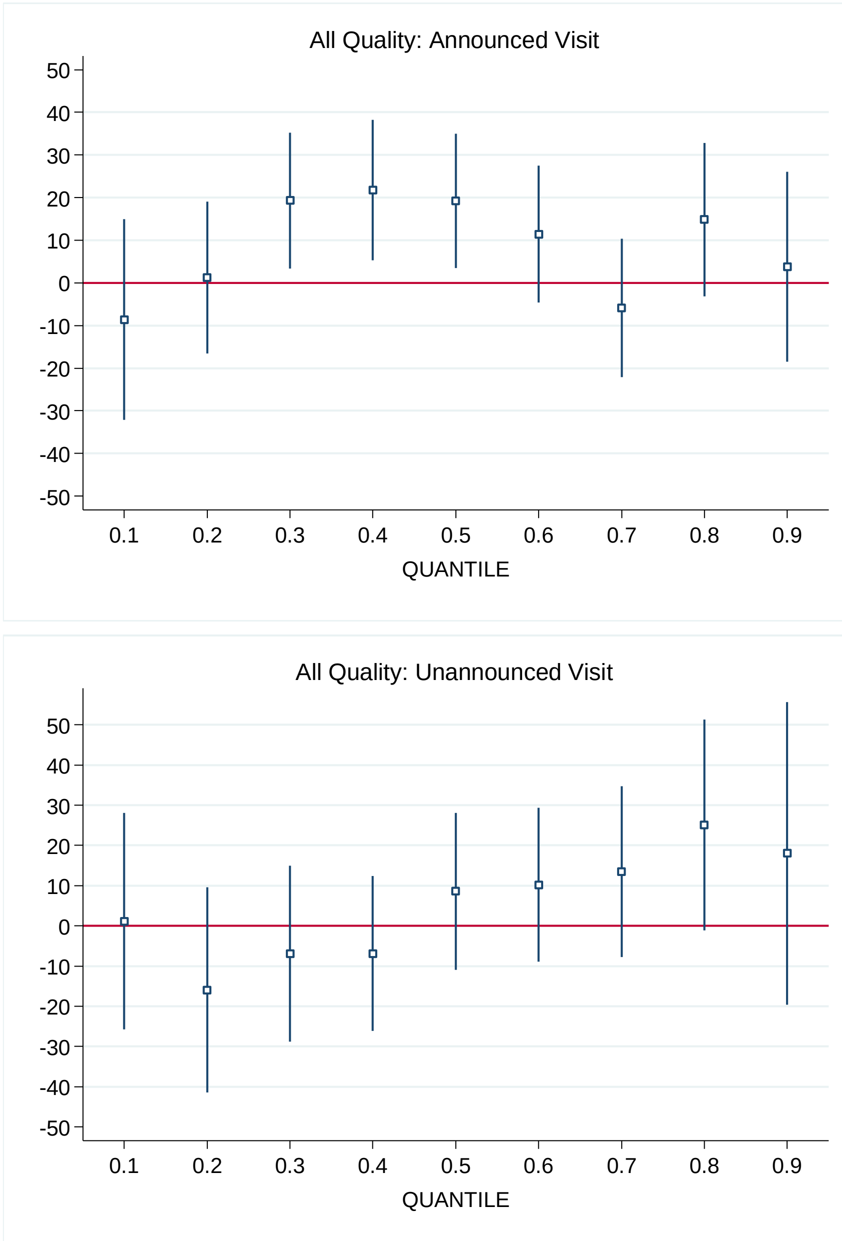
	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
	1	2	3	4	5	6	7	8	9
Announced Visit	-8.626	1.269	19.314*	21.723**	19.223*	11.432	-5.838	14.838	3.742
	(11.982)	(9.044)	(8.101)	(8.395)	(8.034)	(8.168)	(8.258)	(9.149)	(11.331)
DV (*t*-1)	0.540**	0.571**	0.564**	0.575**	0.534**	0.515**	0.524**	0.528**	0.490**
	(0.038)	(0.028)	(0.022)	(0.021)	(0.019)	(0.020)	(0.023)	(0.029)	(0.039)
Constant	179.478**	201.709**	248.203**	274.341**	331.453**	374.899**	416.526**	449.939**	539.208**
	(27.279)	(20.758)	(16.177)	(15.289)	(14.414)	(14.725)	(16.497)	(20.757)	(27.928)
Nursing FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of NHs	552	552	552	552	552	552	552	552	552
Observations	8,280	8,280	8,280	8,280	8,280	8,280	8,280	8,280	8,280
R^2	0.065	0.098	0.121	0.131	0.131	0.128	0.125	0.105	0.069

Cluster-robust standard errors are given in parentheses; ** p<0.01, * p<0.05, †p<0.1.

	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
	1	2	3	4	5	6	7	8	9
Unannounced Visit	1.168	-15.916	-6.904	-6.864	8.581	10.216	13.476	25.083†	18.057
	(13.721)	(12.975)	(11.114)	(9.804)	(9.921)	(9.759)	(10.818)	(13.321)	(19.156)
DV (*t*-1)	0.553**	0.565**	0.533**	0.560**	0.515**	0.513**	0.534**	0.558**	0.544**
	(0.043)	(0.032)	(0.025)	(0.023)	(0.022)	(0.022)	(0.025)	(0.033)	(0.044)
Constant	162.154**	210.177**	271.685**	285.157**	345.669**	378.559**	405.536**	435.687**	503.504**
	(30.456)	(23.150)	(17.968)	(16.808)	(16.348)	(16.744)	(18.960)	(24.178)	(32.231)
Nursing FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of NHs	441	441	441	441	441	441	441	441	441
Observations	6,615	6,615	6,615	6,615	6,615	6,615	6,615	6,615	6,615
R^2	0.076	0.111	0.123	0.136	0.131	0.131	0.123	0.111	0.081

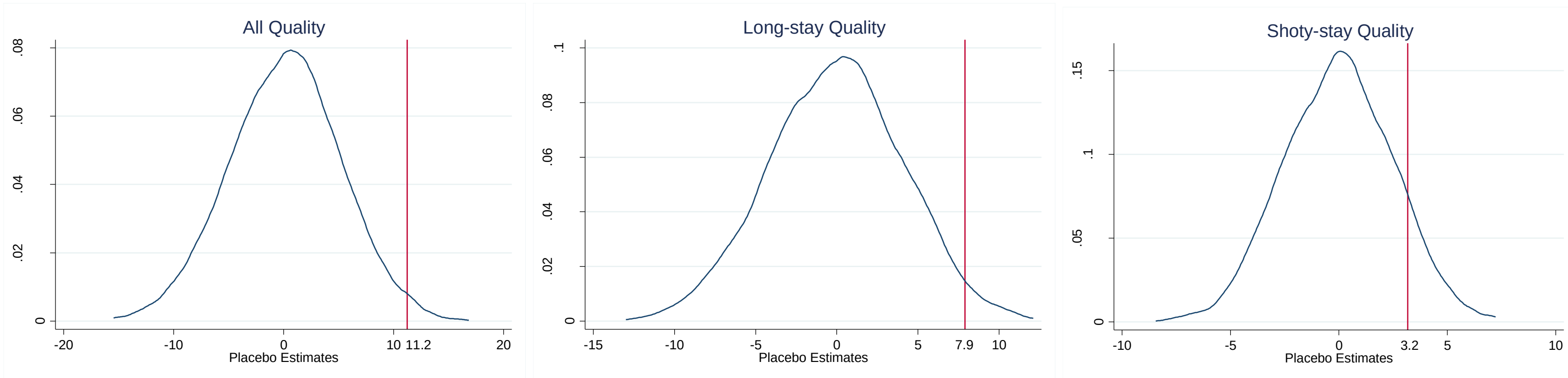
Cluster-robust standard errors are given in parentheses; ** p<0.01, * p<0.05, †p<0.1.

Figure 1 Immediate Effects in Quantile Regression



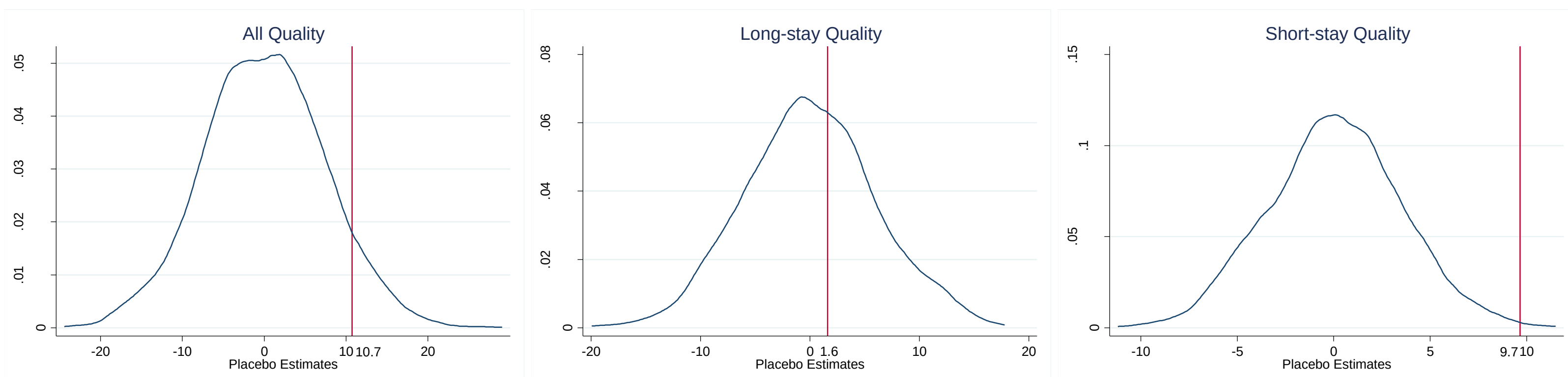
Notes. Dots represent the coefficient estimates of immediate effects with respect to the current standing of quality represented by quantile (decile) levels. Vertical lines indicate 95% confidence intervals for these coefficient estimates.

Figure 2 Kernel Density Plots of Placebo Test Coefficients: Announced Inspection



Notes. The graphs display kernel density plots of the distribution of 2,000 placebo estimates of the effects of the announced inspection visits. Each vertical line indicates the estimate observed in the actual data: 11.2 for all-residents quality, 7.9 for long-stay residents quality, and 3.2 for short-stay residents quality. The graphs except short-stay resident one, indicate that these estimates in actual data are extremely unlikely to occur by chance, which is consistent with our main result.

Figure 3 Kernel Density Plots of Placebo Test Coefficients: Unannounced Inspection



Notes. The graphs display kernel density plots of the distribution of 2,000 placebo estimates of the effects of the unannounced inspection visits. Each vertical line indicates the estimate observed in the actual data: 10.7 for all-residents quality, 1.6 for long-stay residents quality, and 9.7 for short-stay residents quality. The graphs except long-stay resident one, indicate that the estimate in actual data are extremely unlikely to occur by chance, which is consistent with our main result.

Appendix C. Organizational Learning, Complexity, and Risk Management Performance (Chapter 4)

Table 1 Pipeline Operators That Ever-Experienced Incidents and Ever-Conducted BaseAssess

Operators		*Ever-Conducted BaseAssess*		
		Yes	No	Total
Ever-Experienced Incidents	Yes	234**	25	259
	No	230	153	383
	Total	464	178	642*

*Full sample; **Subsample

Observations		*Ever-Conducted BaseAssess*		
		Yes	No	Total
Ever-Experienced Incidents	Yes	2,369**	103	2,472
	No	1,672	552	2,224
	Total	4,041	655	4,696*

*Full sample; **Subsample

Table 2 Robustness Checks: Enforcement as Additional Control

Sample (Complexity)	Sub (All)	Sub (High)	Sub (Low)
DV: *Incident CostPerMile*$_{t+1}$	Model	Model	Model
	1	2	3
Complexity	2.147**	1.099	2.595**
	(0.585)	(1.050)	(0.694)
BaseAssess Indicator	0.566*	0.109	0.780†
	(0.265)	(0.371)	(0.396)
BaseAssess Experience	-0.168**	-0.164*	-0.156
	(0.062)	(0.076)	(0.103)
Enforcement: # of warnings	-0.047	-0.069	0.001
	(0.120)	(0.119)	(0.300)
Enforcement: # of concern letters	0.088	0.034	0.415
	(0.254)	(0.272)	(0.605)
Enforcement: # of amendment notices	-0.075	-0.140	0.097
	(0.110)	(0.105)	(0.278)
Enforcement: # of penalty issued	0.076	0.032	0.360
	(0.134)	(0.140)	(0.390)
Yrs from Last BaseAssess	-0.027	-0.196†	0.056
	(0.052)	(0.100)	(0.061)
Ownership	0.237	0.003	0.430
	(0.303)	(0.295)	(0.547)
HCA Ratio	-0.055	-0.380	0.179
	(0.374)	(0.484)	(0.547)
Pipeline Diameter Large	1.432	2.853	0.288
	(1.755)	(3.232)	(1.802)
Pipeline Diameter Medium	0.263	-0.887	0.768
	(0.917)	(1.526)	(1.160)
Average Age	0.002	-0.019	0.007
	(0.009)	(0.013)	(0.013)
Constant	0.794	3.480**	-0.084
	(0.644)	(1.062)	(0.785)
Pipeline Operator FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
# of Operators	234	117	117
Observations	2,369	1,229	1,140
R^2	0.019	0.023	0.038

Operator-cluster robust standard errors in parentheses; ** p<0.01, * p<0.05, † p<0.1.
Notes. We use a subsample, excluding pipeline operators that never experienced incidents or that never conducted Baseline Assessment. PHMSA measures IMP performance as the number of miles repaired. PHMSA enforce compliance with safety regulations through enforcement actions: warning letters, concern letters, notices of amendment, penalty issued. We additionally control for repair actions, measured the number of actions in each category.

Table 3 Post-hoc Analysis: Hybrid Model Exhibiting With-in and Between Variation

Sample (Complexity)	Sub (All)		Sub (High)		Sub (Low)	
Variation	With-in	Between	With-in	Between	With-in	Between
DV: *Incident CostPerMile*$_{t+1}$	Model 1	Model 1 Cont'd	Model 2	Model 2 Cont'd	Model 3	Model 3 Cont'd
Complexity	2.132**	1.820**	1.149	6.703**	2.504**	0.039
	(0.579)	(0.484)	(1.061)	(1.930)	(0.692)	(0.750)
BaseAssess Indicator	0.573*	-2.430**	0.134	-2.591*	0.794*	-1.660
	(0.264)	(0.855)	(0.369)	(1.183)	(0.393)	(1.074)
BaseAssess Experience	-0.170**	0.695**	-0.173*	0.645**	-0.158	0.615*
	(0.061)	(0.124)	(0.075)	(0.164)	(0.102)	(0.261)
Years from Last BaseAssess	-0.027	-0.234†	-0.196*	0.065	0.050	-0.310*
	(0.051)	(0.125)	(0.097)	(0.231)	(0.060)	(0.138)
Ownership	0.236	-0.081	-0.007	-0.484†	0.373	0.232
	(0.300)	(0.206)	(0.282)	(0.279)	(0.542)	(0.278)
HCA Ratio	-0.075	0.021	-0.117	0.132	0.029	-0.001
	(0.108)	(0.322)	(0.088)	(0.540)	(0.428)	(0.391)
Pipeline Diameter Large	1.435	1.529**	2.830	2.869**	0.304	1.166*
	(1.733)	(0.410)	(3.123)	(0.606)	(1.829)	(0.454)
Pipeline Diameter Medium	0.268	1.120**	-0.871	2.215**	0.672	0.575†
	(0.910)	(0.305)	(1.534)	(0.716)	(1.137)	(0.296)
Average Age	0.002	0.011	-0.017	0.013	0.006	0.013†
	(0.009)	(0.007)	(0.013)	(0.013)	(0.012)	(0.008)
Constant	0.479		-3.196*		0.737	
	(0.539)		(1.456)		(0.605)	
Year FE	Yes		Yes		Yes	
# of Operators	234		117		117	
Observations	2,369		1,229		1,140	
Wald χ^2	304.96		300.64		101.14	

Operator-cluster robust standard errors in parentheses; ** p<0.01, * p<0.05, † p<0.1.

Notes. We use hybrid models, which incorporate within and between effects on *incident consequence* and report operator-cluster robust standard errors in parentheses. We use a subsample, excluding pipeline operators that never experienced incidents or that never conducted Baseline Assessment.